SOCIOPHONETIC VARIATION AT THE INTERSECTION OF GENDER, REGION, AND STYLE IN JAPANESE FEMALE SPEECH

A Dissertation submitted to the Faculty of the Graduate School of Arts and Sciences of Georgetown University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Linguistics

By

Sakiko Kajino, M.S.

Washington, D.C.
March 18th, 2014
SOCIOPHONETIC VARIATION AT THE INTERSECTION OF GENDER, REGION, AND STYLE
IN JAPANESE FEMALE SPEECH
Sakiko Kajino, M.S.

Dissertation Advisors: Natalie Schilling, Ph.D. and Robert J. Podesva, Ph.D.

ABSTRACT

This dissertation is a sociophonetic study of 46 female Japanese speakers from three major metropolitan regions: Tokyo, Kyoto, and Osaka. While previous work on Japanese Women’s Language assumes a monolithic speech variety, this study shows that women in the three regions exhibit strikingly different speech patterns. Rather than constructing a uniform gender identity, Japanese women produce gendered figures that typify particular geographic regions while negotiating the regional stereotypes.

Three phonetic features in 25 dyadic conversation recordings of 46 participants are analyzed quantitatively and qualitatively: breathy voice, acoustic characteristics of voiceless sibilant fricatives /s/ (e.g. sumi ‘charcoal’) and /ɕ/ (e.g. shumi ‘hobby’), and intonational patterns (accented vs. deaccented) of negative polar questions (e.g. amakunai? ‘isn’t [this] sweet?’). The analyses present the cross-regional patterning as well as intra-regional variation using the mixed-method technique with sociolinguistic variationist analysis, close examination of conversations, and ethnographic approach.

The cross-regional analyses, which present big-picture patterns for the three phonetic features, show the following:
1) A feature that is considered to mark gender (i.e. breathy voice) exhibits regional differences (for Kyoto speakers, breathy voice exhibits a stronger correlation with low intensity and high F0 levels than for Tokyo and Osaka speakers)
2) A feature that serves to distinguish region (acoustic analyses of the fricatives /s/ and /ɕ/ show that the Tokyo fricatives are significantly different from the Osaka fricatives) simultaneously connotes meanings that can be used to construct gender (e.g. higher center of gravity of fricatives connotes “sharpness”); and

3) A feature that carries the meaning of the Tokyo regionality (i.e. the deaccented form of negative questions) can be used by speakers of other regions to indicate their alignment with a Tokyo-centric ideology.

Intra-regional variation is further examined to explore the meanings of the quantitative patterns at the interactional level. The meanings are drawn based on the participants’ individual styles that are co-constructed by linguistic and non-linguistic identity practices. The intra-regional analyses reveal how the participants utilize the phonetic features to construct their regional gender identities while aligning with or disaligning from the local stereotypes, such as the “boring” Tokyo, the “classy” Kyoto, and “harsh” Osaka.
ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of all the people I met during my journey at Georgetown. First of all, I would like to thank the four members of my dissertation committee: Rob Podesva, Natalie Schilling, Shigeto Kawahara, and Mark Sicoli. Thank you for your wise guidance, thoughtful criticism, continuous encouragement, and the countless hours you spent helping me. Especially, I owe Rob and Natalie so much not only for being exceptionally compassionate mentors but also for changing my life completely; without them, I would never have thought that I could pursue a PhD. Natalie, you were the first professor in my life who made me think that my ideas can be worthwhile. The very first paper I wrote in your Intro to Socio class brought me all the way here, and this dissertation is built on the most important lessons you taught me in sociolinguistics. Rob, I am extremely fortunate to have worked with someone like you, someone with whom I can share my enthusiasm. Your encouragement to be creative and innovative shaped my identity as a linguist, and this dissertation is the outcome of the challenges and inspiration you offered me all these years. It was also such an honor to have Shigeto and Mark in my committee. Throughout my graduate research, whenever I encountered questions on Japanese phonetics and phonology, Shigeto’s work provided me with the answers. Shigeto brought me completely new cross-disciplinary perspectives, and his input was tremendous and essential in this dissertation. Mark was extraordinary generous with his very-last-minute help, and regardless of the tight schedule, his comments were very precise and thought-provoking. His scholarship motivates me greatly, and one day, I hope to produce work like his.

Second, I would like to thank the UCLA phoneticians who let me participate in their data analysis sessions, phonetic seminars, and class. Particularly, I want thank Sun-Ah Jun, who let me audit her *Intonation* class, and Pat Keating and Megha Sundara, who were very
understanding and provided me with an opportunity to talk about the early findings of this dissertation. I also want to thank the sociolinguists at the National Institute for Japanese Language and Linguistics (NINJAL), in particular, Yoshiyuki Asahi, who let me talk about the preliminary study of this dissertation at a colloquium, and Janet Stuart-Smith and Ichiro Ohta, who gave me very valuable comments at the colloquium. Thank you, Akira Satoh and his students at Osaka University, for letting me talk about the project proposal in your seminar and sharing very insightful thoughts about the Osaka variety.

Third, I am deeply indebted to the three anonymous professors at the three women’s colleges for providing me with such great assistance for my fieldwork. Also to 46+ participants of this study, thank you for sharing your time and pieces of your life with me, and sometimes showing me your support and encouragement through letters and text messages. I wish you the best in your lives, and this dissertation is dedicated to all of you.

To my precious friends, especially Chie Adachi, Asuka Goto, Atsuko Koyama, Mika Hama, Hirono Ishigami, Emily Klonicki, Sinae Lee, Kyuwon Moon, Bokyun Mun, Satoshi Nambu, Anastasia Nylund, Anna Strycharz-Banas, Motoko Takada, Ayumi Tokunaga, Robin Rockey, Christoph Ruesch, Matt Wieland, and so many others; thank you for being there and helping me through this.

To my parents, Masatoshi and Junko Kajino and Bill and Reenie Murley, thank you for letting me live a life filled with love.

Finally, to C – thank you. This is our baby, and we will soon have more
# TABLE OF CONTENTS

Chapter 1. Introduction: Base of the Study................................................................. 1

1.1. Overview............................................................................................................. 1
   1.1.1. Research Questions .................................................................................. 1
   1.1.2. Linguistic Variables ................................................................................ 2
   1.1.3. Data and Participants ............................................................................. 2
   1.1.4. Methodological and Theoretical Approaches....................................... 3
   1.1.5. Terminological Excursions .................................................................... 4

1.2. Overview of the Tokyo, Kyoto, and Osaka Language Varieties ................. 6

1.3. Relevant Field I: Japanese Dialectology ......................................................... 11
   1.3.1. Standardization of Japanese ................................................................ 11
   1.3.2. Post-Standardization ............................................................................. 15
   1.3.3. Research on Japanese Dialectology ..................................................... 17

1.4. “Standard” Varieties and Gender in Western Communities ...................... 19

1.5. Relevant Field II: Japanese Women’s Language ........................................... 23
   1.5.1. Socio-Historical Context of Japanese Women’s Language ................. 24
   1.5.2. Research on Japanese Women’s Language ......................................... 28
   1.5.3. Japanese Women’s Language and Region ......................................... 30

1.6. Participants’ Attitudes toward Regional Varieties and JWL....................... 32
   1.6.1. Feminine and Masculine Ratings ......................................................... 32
   1.6.2. Metalinguistic Commentaries and Local Stereotypes ....................... 38
   1.6.3. Self-Evaluations of Regional Speech ............................................... 43
   1.6.4. Metalinguistic Commentaries and Japanese Women’s Language .... 46

1.7. Current Study................................................................................................... 49
   1.7.1. General Research Questions of this Dissertation ......................... 49
   1.7.2. Linguistic Features .............................................................................. 50

1.8. Social Meaning ............................................................................................... 52
   1.8.1. Three Waves of Sociolinguistic Variation ....................................... 52
   1.8.2. Indexical Meaning ............................................................................ 54
   1.8.3. Indexical Field ................................................................................ 59
   1.8.4. Definition of Social Meaning ......................................................... 59

1.9. Summary........................................................................................................... 61

1.10. Organization of the Dissertation ................................................................. 61

Chapter 2. Fieldwork............................................................................................... 64

2.1. Introduction....................................................................................................... 64

2.2. Research Sites ............................................................................................... 64
2.2.1. Rationale ..............................................................................64
2.2.2. Tokyo Rose, Kyoto Cherry, and Osaka Lily .......................67
2.3. Initial Contact with the Research Sites ................................70
2.4. Mini-Ethnography .................................................................71
2.5. Participants in Audio Recordings ..........................................74
2.6. Recordings of Conversations .................................................75
2.7. Interviews ............................................................................79

Chapter 3. Breathy Voice: The Essence of Elegance ...................82
3.1. Introduction ...........................................................................82
3.2. Background I: Phonetics of Breathy Voice .........................82
   3.2.1. Phonation ......................................................................82
          3.2.1.1. Breathy Voice vs. Whispherly Voice ..................84
   3.2.2. Acoustic Correlates of Breathy .....................................85
          3.2.2.1. Spectral Measures ............................................85
          3.2.2.2. Periodicity .....................................................88
3.3. Background II: Breathy Voice and Gender .........................89
   3.3.1. Male-Female Differences ............................................89
          3.3.1.1. Non-Physiological Based Differences ...............91
   3.3.2. Perception Studies ...................................................92
   3.3.3. Age ..........................................................................93
   3.3.4. Dialects .......................................................................94
3.4. Current Study .......................................................................94
   3.4.1. Introduction ................................................................94
   3.4.2. Procedure ..................................................................95
          3.4.2.1. Additional Linguistic Features .......................98
   3.4.3. Results .......................................................................101
   3.4.4. Discussion ..................................................................108
3.5. Conclusion ..........................................................................110

Chapter 4. Fricatives: The Source of Noise ...............................112
4.1. Introduction .........................................................................112
4.2. Background I: Phonetics of Sibilant Fricatives ....................113
   4.2.1. Voiceless Sibilant Fricatives .....................................113
   4.2.2. English Fricatives vs. Japanese Fricatives Articulation ....115
   4.2.3. Measurements of Fricatives .......................................116
          4.2.3.1. Amplitude of Noise .......................................117
          4.2.3.2. Noise Duration .............................................118
4.2.3.3. Spectral Analysis

4.3. Background II: Fricatives and Gender
4.3.1. Male-Female Differences
4.3.2. Non-Physiological Patterns
4.3.3. Age
4.3.4. Trans Identities
4.3.5. Cross-linguistic Cases
4.3.6. Japanese Fricatives and Gender

4.4. Psychoacoustic Parameters of Sound Image

4.5. Current Study
4.5.1. Introduction
4.5.2. Procedure
4.5.3. Results
4.5.4. Discussion

4.6. Conclusion

Chapter 5. Deaccented Intonation: The Spice of Locality
5.1. Introduction
5.2. Background I: Japanese Intonational Systems
5.2.1. Tokyo Japanese Prosodic Organization
5.2.2. Map of Japanese Regional Intonational Patterns
5.2.3. Japanese Dialectology and Intonation
5.3. Background II: Deaccented Negative Polar Questions
5.3.1. Deaccented Negative Polar Questions in Japanese
5.3.2. Meaning of Rising Intonation
5.3.3. Meaning of Negative Polar Questions
5.3.4. Meaning of Japanese Negative Polar Questions
5.3.5. Tonal Compression of Japanese Negative Polar Questions
5.3.6. Meaning of Tonal Reduction
5.3.7. Meaning of Deaccentuation
5.4. Current Study
5.4.1. Introduction
5.4.2. Procedure
5.4.3. Results
5.4.4. Discussion
5.4.4.1. Grammatical Categories
5.4.4.2. Pattern in Tokyo
5.4.4.3. Pattern in Kyoto
5.4.4.4. Pattern in Osaka
5.5. Conclusion

Chapter 6. Style and Meaning: The Recipe for Femininities
Chapter 6. Conclusion: Food for Thought

7.1. Overview .................................................................................. 248

7.2. Chapter Summary ..................................................................... 249

7.3. Contributions ........................................................................... 253

7.4. Limitations and Future Studies .................................................. 255

Appendix I: Participants List .............................................................. 260

Appendix II: Interview Questionnaire ................................................ 261

Appendix III: Original Japanese Transcriptions of Excerpts in Chapter 6 .................................................. 262

References .......................................................................................... 268
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Dialectal divisions of Japan based on cluster analysis of “Standard Japanese” forms (F. Inoue 1995: 358, modified by the author)</td>
<td>8</td>
</tr>
<tr>
<td>1.2</td>
<td>Femininity ratings among the three varieties</td>
<td>35</td>
</tr>
<tr>
<td>1.3</td>
<td>Masculinity ratings among the three varieties</td>
<td>37</td>
</tr>
<tr>
<td>1.4</td>
<td>Word cloud for the image of the Tokyo variety</td>
<td>39</td>
</tr>
<tr>
<td>1.5</td>
<td>Word cloud for the image of the Kyoto variety</td>
<td>40</td>
</tr>
<tr>
<td>1.6</td>
<td>Word cloud for the image of the Osaka variety</td>
<td>40</td>
</tr>
<tr>
<td>1.7</td>
<td>Osaka Auntie illustrations on the book cover of Osaka no Obachan Korekushon [Osaka Auntie Collections] (Maegaki 2014)</td>
<td>42</td>
</tr>
<tr>
<td>1.8</td>
<td>Word cloud for the image of Japanese Women’s Language</td>
<td>48</td>
</tr>
<tr>
<td>2.1</td>
<td>Examples of female and male fashion magazines used during the recordings</td>
<td>78</td>
</tr>
<tr>
<td>3.1</td>
<td>Midsagittal section of the vocal tract (Kent and Read 2002: 9)</td>
<td>83</td>
</tr>
<tr>
<td>3.2</td>
<td>States of the glottis according to phonatory settings (Catford 1988: 54)</td>
<td>84</td>
</tr>
<tr>
<td>3.3</td>
<td>FFT spectra for modal voice and breathy voice of the vowel [a] taken from a Tokyo speaker</td>
<td>86</td>
</tr>
<tr>
<td>3.4</td>
<td>Box plots of the measurements for breathy voice</td>
<td>105</td>
</tr>
<tr>
<td>3.5</td>
<td>Scatter plots of H1-H2 and intensity ratio across three regions</td>
<td>106</td>
</tr>
<tr>
<td>3.6</td>
<td>Scatter plots of H1-H2 and F0 across three regions</td>
<td>107</td>
</tr>
<tr>
<td>3.7</td>
<td>Midsagittal sections during the production of the two fricatives (Stevens 1998: 380)</td>
<td>114</td>
</tr>
<tr>
<td>4.1</td>
<td>FFT spectrum of [s] in “sonna (‘such’)” produced by a Tokyo speaker</td>
<td>118</td>
</tr>
<tr>
<td>4.2</td>
<td>Box plots of the measurements for /s/</td>
<td>137</td>
</tr>
<tr>
<td>4.3</td>
<td>Box plots of the measurements for /ɕ/</td>
<td>140</td>
</tr>
<tr>
<td>5.1</td>
<td>Conventional accented pattern (Adapted from Hara and Kawahara 2012)</td>
<td>145</td>
</tr>
<tr>
<td>5.2</td>
<td>New deaccented pattern (Adapted from Hara and Kawahara 2012)</td>
<td>146</td>
</tr>
<tr>
<td>5.3</td>
<td>Fo contour and J-TOBI transcription of the utterance</td>
<td>151</td>
</tr>
<tr>
<td>5.4</td>
<td>Map of intonational patterns of Japan (Adapted from Kindaichi 2001)</td>
<td>154</td>
</tr>
<tr>
<td>5.5</td>
<td>Examples of bimoraic words in the Tokyo style pattern and Kyoto-Osaka style pattern (Adapted from Fujisaki et al. 1976: 174)</td>
<td>157</td>
</tr>
<tr>
<td>5.6</td>
<td>Example of tonal compression in the utterance of Example 12 (Adapted from Ito and Oshima 2013: 3)</td>
<td>172</td>
</tr>
</tbody>
</table>
Figure 5.7. Example of retained accent in the utterance of Example 13 (Adapted from Ito and Oshima 2013: 3) .................................................................................................................. 172
Figure 5.8. F0 contours in the utterance kyo’nen ro’omani ikima’shita ‘[I] went to ROME last year’ (Adapted from Venditti, Maekawa, and Beckman 2008: 485).............................. 174
Figure 5.9. Distributions of the deaccented/accented forms across the regions........ 179
Figure 5.10. One Piece characters on Orie’s backpack .......................................................... 195
Figure 5.11. Orie’s Evangelion wallet.................................................................................. 195
Figure 6.1. Inter-speaker variation in H1-H2 and COG of /s/ .............................................. 213
Figure 6.2. Sample advertisement of Jun’s favorite fashion brand ...................... 219
Figure 6.3. Sample advertisement of Mei’s favorite fashion brand ...................... 219
LIST OF TABLES

Table 2.1. Summary of Research Sites ................................................................. 68
Table 3.1. Mixed effects for model of H1-H2 ......................................................... 102
Table 3.2. Mixed effects for model of H1-A1 ......................................................... 103
Table 3.3. Mixed effects for model of H1-A2 ......................................................... 103
Table 3.4. Mixed effects for model of H1-A3 ......................................................... 104
Table 4.1. Mixed effects for model of Spectral Peaks (ERB) for /s/ ......................... 135
Table 4.2. Mixed effects for model of COG (ERB) for /s/ ...................................... 136
Table 4.3. Mixed effects for model of SD for /s/ ................................................... 136
Table 4.4. Mixed effects for model of Skewness for /s/ ......................................... 136
Table 4.5. Mixed effects for model of Kurtosis for /s/ ............................................ 136
Table 4.6. Mixed effects for model of Intensity for /s/ ........................................... 137
Table 4.7. Mixed effects for model of Spectral Peaks (ERB) for /ɕ/ ......................... 138
Table 4.8. Mixed effects for model of COG (ERB) for /ɕ/ ...................................... 138
Table 4.9. Mixed effects for model of SD for /ɕ/ ................................................... 138
Table 4.10. Mixed effects for model of Skewness for /ɕ/ ........................................ 138
Table 4.11. Mixed effects for model of Kurtosis for /ɕ/ .......................................... 138
Table 4.12. Mixed effects for model of Intensity for /ɕ/ ......................................... 138
Table 4.13. Summary of Results for Regions ....................................................... 141
Table 5.1. Tokyo accentual patterns .................................................................... 156
Table 5.2. Kyoto-Osaka accentual patterns ......................................................... 156
Table 5.3. Summary of conclusions of Büring and Gunlogson (2001) .................. 167
Table 5.4. Summary of Japanese NPQs and epistemic and evidential biases (Created based on Sudo 2013) .............................................................. 169
Table 5.5. Mixed effects for model of deaccentuation ......................................... 178
Table 6.1. Summary of individual means and ranks across linguistic features ........ 216
Table 6.2. Summary of inter-speaker variation between Jun and Mei from Tokyo .... 218
Table 6.3. Summary of inter-speaker variation between Miya and Nanako from Osaka ................................................................. 226
Table 6.4. Summary of inter-speaker variation between Eri and Mika from Kyoto .... 226
Table 6.5. Summary of inter-speaker variation across participants with high degrees of breathy voice ....................................................................................... 229
Table 6.6. Summary of inter-speaker variation between Sae and Mami from Kyoto .. 235
Table 6.7. Summary of inter-speaker variation between Sayumi and Orie from Kyoto..... 237
Table 6.8. Summary of inter-speaker variation between Chie and Futaba from Kyoto..... 238
Table 6.9. Summary of inter-speaker variation between two Osaka participants........... 240
Table 6.10. Summary of inter-speaker variation between four Osaka participants ........ 243
CHAPTER 1

Introduction: Base of the Study

1.1. Overview

This dissertation is a sociophonetic study of female Japanese speakers from three major metropolitan regions: Tokyo, Kyoto, and Osaka. Through examinations of the language varieties of Tokyo, Kyoto, and Osaka, this study relates the discussion to the fields of Japanese Women’s Language (JWL) and Japanese Dialectology. While previous work on Japanese Women’s Language assumes a monolithic speech variety, this dissertation shows that women in Tokyo, Kyoto, and Osaka exhibit strikingly different speech patterns. Rather than constructing a uniform gender identity, Japanese women produce gendered figures that typify particular geographic regions while negotiating the local stereotypes. This work problematizes previous work on both Japanese Women’s Language – which homogenizes the diversity of linguistic practices among Japanese women from different regions – and Japanese Dialectology – which locates Osaka and Kyoto within the same dialect region.

1.1.1. Research Questions

This dissertation attempts to elucidate the construction of femininities by the female participants from different regions. Three specific research questions will be listed in Section 1.7.1, after relevant issues on Japanese Dialectology, Japanese gendered language ideology, and local stereotypes of the three regions are addressed.
1.1.2. Linguistic Variables

The three phonetic features examined in this dissertation are *breathy voice*, *acoustic characteristics of voiceless sibilant fricatives* /s/ and /ɕ/, and *intonation of negative polar questions*. Through examination of these features, I will investigate how they pattern among female speakers from the Tokyo, Kyoto, and Osaka regions. While breathy voice and fricatives are analyzed as continuous variables varying along such dimensions as degree of breathiness and height of frequency energy of the fricatives, the intonation of negative polar questions is analyzed in terms of two categorical variants, *accented* questions and *deaccented* questions. More details about each feature as well as rationale for selecting these three phonetic features will be provided in Section 1.7.2, after I review different linguistic features investigated in previous literature on Japanese Dialectology and Japanese women’s linguistic practices.

1.1.3. Data and Participants

There are three sets of data from which I am drawing my analysis in this study: recorded conversations, responses to interview questions, and detailed field notes from a mini-ethnography. The main data on which the linguistic investigations of the three phonetic variables are based consists of over 40 hours of 25 recordings of conversations with 46 female college students between the ages of 18 and 21 years old. The 46 participants include 15 students from Tokyo, 15 students from Kyoto, and 16 students from Osaka, who were all recruited on the campuses of Tokyo Rose University, Kyoto Cherry University, and Osaka Lily University (pseudonyms). The spontaneous conversations were recorded with pairs of students without the researcher’s presence. In addition to these recorded conversations, I conducted interviews with each pair of participants focusing on their demographic background and beliefs and ideologies about language and gender. A mini-ethnography was also conducted in order to become familiar
with the participants’ communities of practice. Details about the data, participants, and fieldwork will be discussed in Chapter 2.

1.1.4. Methodological and Theoretical Approaches

This study presents mixed-methods research utilizing both quantitative and qualitative sociolinguistic approaches. Chapters 3, 4, and 5 present results of sociolinguistic variationist analyses to uncover the linguistic and extralinguistic patterning of breathy voice, fricatives, and intonation of negative polar questions among the participants from the Tokyo, Kyoto, and Osaka regions. The distributions of these linguistic features across the three regions and correlations with a number of linguistic factors are analyzed with mixed effects regression modeling with the statistical software R. Results reveal how the speech patterns of women from one region are different or similar to the patterns of women from the other regions. Later in this chapter, I will outline the main methodological and analytic approaches used in sociolinguistic variationist studies as well as in my study.

While quantitative analyses provide a big picture of language variation, they do not always reveal in detail what the variants mean to the speakers who use them. Therefore, in Chapter 5, I perform in-depth qualitative analysis in addition to the statistical analysis of cross-regional distributions of the accented or deaccented forms of negative questions. Close examination of how the deaccented form is used in unfolding discourse highlights the locally specific meaning of the feature, which is not evident from examining the quantitative patterns alone.

Chapter 6 utilizes an ethnographic approach to explore the meaning of variation in breathy voice and acoustic characteristics of fricatives as well as the meaning of linguistic styles comprised of various combinations of these and other linguistic features. Following Eckert
(1996), Bucholtz (1999), and Johnstone (1995), I consider features of linguistic style to cluster not only with other co-occurring linguistic features but also with non-linguistic features of stylistic practices. Language styles are not detachable from non-linguistic practices such as clothing and actions since linguistic and non-linguistic factors have mutual reinforcing effects to construct a style. By providing information on the participants’ non-linguistic practices, an ethnographic approach helps us understand the meanings of language styles as well as variation.

1.1.5. Terminological Excursions

**Sex and Gender**

The commonly used distinction between *sex* and *gender* essentially recognizes biological and sociocultural differences. Sex refers to the biological/bodily classification of living beings as female and male, while gender refers to sociocultural practices, conventions and ideologies clustering around the biological classification (McConnell-Ginet 2011: 6). This theoretically sharp demarcation falls short in reality since there is no single purely biological criterion for male or female sex. Sex is based in a combination of anatomical, endocrinal, and chromosomal features, and the selection among these criteria for sex assignment is based very much on cultural beliefs about what actually makes someone male or female (Eckert and McConnell-Ginet 2013: 2). Bing and Bergvall (1996) describe that in most cultures, medical intervention polices the boundaries to ensure that newborn babies fit into the binary physiological categories, but there are other cultures where more than two categories (or an ambiguous category) are given explicit social recognition.

In sociolinguistics, issues of gender emerged primarily as the study of “sex differences” by focusing on the quantifiable difference between women’s and men’s use of particular linguistic variables, but scholars soon reached the realization that very few biological differences
have straightforward effects on language. Cameron (2009) points out that those pushing for biologically based explanations of sex differences overlook the fact that the very same linguistic differences that they see between the sex categories also correlate with other social categories such as class and race, and many of the “sex differences” they cite as biologically based actually vary historically and cross-culturally. For example, the difference in an individual’s voice pitch is one of the few features believed to be rooted in his or her physiology, but studies (e.g. Ohara 1992, 1997, Van Bezooijen 1995, 1996, Yuasa 2010, Stanford 2010) have shown that this is not necessarily the case.

Nevertheless, sociolinguists have continued to examine the interaction between gender and variation by correlating variables with sex rather than gender differences since information about a speaker’s sex is easily accessible, but his or her gender-related place in society is “a multidimensional complex that can only be characterized through careful analysis” (Eckert 1989: 247). Chambers (2003: 118-19) distinguishes sex-based variability and gender-based variability, noting that in the case of the former, male-female differences will persist even in the absence of well-defined gender roles in a given community, whereas the latter require differentiation of gender roles. However, as Cheshire (2002) notes, it is very difficult to keep the two concepts apart, especially among studies that were designed with a categorization of individuals by their sex but that are then interpreted in terms of social aspects of gender.

The term gender, as I use it in my study, encompasses what is conventionally treated as sex (biological/physical) and gender (sociocultural) even in discussion of sociolinguistic work that uses the label sex. Considering these competing and conflicting definitions of sex and gender, this study adopts the alternate definition of gender proposed by Bucholtz (2002: 37) that captures more fully the social grounding of both sex and gender: “The negotiable and contestable social classification of individuals into the categories of female and male based on cultural understandings of the body, especially with respect to sexuality, with attendant
normative local ideologies about social, physical, cognitive, and affective practices, attributes, and capabilities.” In addition to the fact that the biological/physical aspect of femaleness or maleness is shaped by social ideologies (Zimman 2010), I regard much gender display to be sexual or sexualized display since sexuality, another concept central to individual experience and social life, is intricately intertwined with gender.

**Femininity**

Throughout the study, I discuss multiple femininities displayed by the participants through their linguistic and non-linguistic practices. By using the term femininities, I do not mean to discuss different degrees of “womanliness”; rather, I am concerned with different women’s personas or identities and different ways to be women. Some of the participants in this study appear to disengage in or even reject what is commonly considered as “typical” female practice, but this is still a way to construct a femininity. Identities are mutually constructed by speakers and listeners who interpret and comprehend speech, appearance, and actions. Therefore, as long as a speaker is seen as a female, her gender performance or performativity is a display of a femininity.

**1.2. Overview of the Tokyo, Kyoto, and Osaka Language Varieties**

This chapter illuminates the three regions’ dialectal images and provides overviews of two relevant fields: Japanese Women’s Language and Japanese Dialectology. Central theoretical and methodological approaches and key notions of this dissertation will be discussed toward the end of this chapter.

Geographically, Tokyo is located in the eastern part of Japan, while Osaka and Kyoto are located in the western part of Japan. The distance between Tokyo Prefecture and Osaka
Prefecture is approximately 400 km, which is slightly longer than the distance between Washington, D.C. and New York City. Tokyo and Osaka are connected by a fast bullet train, and there is heavy automobile traffic between the two cities. Kyoto Prefecture is located 40 km north east of Osaka Prefecture, and Kyoto and Osaka are connected by subway systems within half an hour of each other. Osaka and Kyoto, together with Kobe, constitute the Greater Osaka Area, commonly called Keihanshin, just as Tokyo, Yokohama, Kawasaki and other suburban cities comprise the Greater Tokyo Area in the East. These two regions are the largest metropolitan areas in Japan. Tokyo has been the capital of Japan since the end of the 19th century, but until then, Kyoto remained the seat of Japan's imperial court for over a thousand years. Osaka has been and continues to be highly commercially active, and it is Japan's second largest city today.

Though Japan is a geographically small country, rich regional linguistic variation is observed. No single map of dialect divisions is agreed upon, but there is a rough consensus among dialectologists (e.g. Kokugo Chōsaiinkai 1906, Tojo 1927, 1954, Tokugawa 1993) as well as the general public (see F. Inoue 1991, 1995) that the major divisions of great significance are between the Okinawa dialect and the mainland dialects, and within the mainland, between the Western dialect group and the Eastern dialect group. Each of the Western and Eastern groups is further divided into three or four smaller dialects. As Figure 1.1 shows, Tokyo is located in the Kanto region of Eastern Japan, and Kyoto and Osaka are both in the Kinki region of Western Japan.
Regional differences between the Eastern dialects including Tokyo and the Western dialects including Kyoto and Osaka are observed in various linguistic features such as the lexicon, phonetic and phonological features, morpho-syntactic structures, and intonational systems. For example, Tojo (1955: 55) provides a list of sample lexical differences between the Tokyo dialect and the Kyoto-Osaka dialect, which contains the following items:
Regarding phonetic variation, Shibatani (1990) notes that while the high back vowel /u/ is produced with unrounded lips in Tokyo, speakers of Kyoto-Osaka Japanese tend to pronounce it more rounded. Also, in the Kyoto-Osaka dialect, monomoraic words1 in Tokyo such as me ‘eye’ and ki ‘tree’ tend to be elongated with prosodic lengthening as mee and kii (Labrune 2012: 253).

Regarding morpho-syntactic variation, verbal inflectional patterns for negative forms present an outstanding contrast between Tokyo and Kyoto-Osaka, as in the following:

<table>
<thead>
<tr>
<th>Plain</th>
<th>English Gloss</th>
<th>Tokyo negative</th>
<th>Kyoto-Osaka negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iku</td>
<td>‘to go’</td>
<td>Ikanai (‘not go’)</td>
<td>Ikahen / Ikehen / Ikan</td>
</tr>
<tr>
<td>Aru</td>
<td>‘to be / to exist’</td>
<td>Nai (‘not exist’)</td>
<td>Arahen / Arehen</td>
</tr>
<tr>
<td>Suru</td>
<td>‘to do’</td>
<td>Shinai (‘not do’)</td>
<td>Seehen / Sen</td>
</tr>
<tr>
<td>Kuru</td>
<td>‘to come’</td>
<td>Konai (‘not come’)</td>
<td>Koohen / Kon</td>
</tr>
</tbody>
</table>

(based on Ueno 2003)

---

1 Moras are the phonological units of timing in Japanese and are traditionally considered to be sub-units of syllables that consist of one short vowel and any preceding onset consonants. Some basics of Japanese prosody will be reviewed in Chapter 5.
Finally, though the intonational systems in Tokyo and Kyoto-Osaka as well as Japanese prosody will be reviewed at length in Chapter 5, I will list a few examples of lexical accentual/pitch patterns below. In the list, ‘L’ indicates low pitch and ‘H’ indicates high pitch.

<table>
<thead>
<tr>
<th>Word</th>
<th>English Gloss</th>
<th>Tokyo pattern</th>
<th>Kyoto-Osaka pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sora</td>
<td>‘sky’</td>
<td>HL</td>
<td>LH</td>
</tr>
<tr>
<td>Miyako</td>
<td>‘capital’</td>
<td>LHH</td>
<td>HHH</td>
</tr>
<tr>
<td>Atama</td>
<td>‘head’</td>
<td>LHH</td>
<td>HLL</td>
</tr>
<tr>
<td>Bentoo</td>
<td>‘lunch box’</td>
<td>LHHL</td>
<td>HHLL</td>
</tr>
<tr>
<td>Kaminari</td>
<td>‘thunder’</td>
<td>LHHH</td>
<td>LHHL</td>
</tr>
</tbody>
</table>

The variation described in this section is based on the varieties spoken in Tokyo and Kyoto-Osaka today. Though a wide range of regional differences still remain, all Japanese language varieties including the Tokyo and Kyoto-Osaka varieties underwent the strict governmental policy of language standardization during Japan’s modernization period (1868-World War II). In the following section, I will review the history of Japanese dialects by highlighting the national standardization efforts through which a new linguistic variety was created based on a specific type of the (male) Tokyo variety, which came to be considered the “standard language.” As J. Milroy (2001) emphasizes, in variationist sociolinguistic studies of what is regarded as “standard” and “non-standard” language varieties, it is important to consider the historical process of language standardization.

Previous literature has traced processes of national language standardization in different places at different times, including France during the 18th century (Parisian French in Bourdieu
1991), England during the 18\textsuperscript{th}-19\textsuperscript{th} centuries (Received Pronunciation (RP) in Agha 2003, 2007), and China during the first half of the 20\textsuperscript{th} century (Putonghua in Dong 2010). Just as was the case in Japan during the 19\textsuperscript{th}-20\textsuperscript{th} centuries, the French, English, and Chinese processes took place in parallel with modern nation building. The primary difference between those other countries’ and Japanese movements is that, as we will see in Section 1.5, the Japanese standardization brought women and men far more different experiences. In Japan, the “standard language” for men was institutionally imposed, as with Putonghua in China (Dong 2010), while the “standard language” for women was propagated in large part through popular print media, as with RP in England (Agha 2003).

1.3. Relevant Field I: Japanese Dialectology

1.3.1. Standardization of Japanese

The existence of dialects in Japan was documented as early as the 8\textsuperscript{th} century in Manyōshū, the oldest existing collection of Japanese poetry. Songs like asuma-uta ‘eastern songs’ and sakimori-uta ‘songs of the garrison soldiers’ are written in the eastern dialect, which was distinct from the variety of western Japan. Thought it is not clear if there was any concept similar to the “standard” language in 8\textsuperscript{th} century Japan, the Kyoto variety became the de facto “standard” Japanese by the early 17\textsuperscript{th} century. A Christian missionary, Rodrigues (1604-08), noted that the language that was used by the court nobles of the then capital, Kyoto approximated something we call “standard Japanese” today (Shibatani 1990: 185). Additionally, in the glossaries of local dialects that were written during this time period, other regional expressions were translated into their Kyoto equivalents.

Even after the main government was transferred to Edo (presently Tokyo) in 1603, Kyoto remained Japan’s capital until 1869 at the time of the Imperial Restoration. Kyoto and its
neighboring city of Osaka comprised a cultural and economic center and dominated the rest of Japan. Therefore, the de facto “standard” status of the Kyoto dialect did not decline very quickly. Hepburn (1872: 14; cited in Lee 2010: 47), who completed a Japanese-English dictionary, reported, “the language of Kyoto, the capital where the Emperor and educated people live, is considered as the most authoritative, standard language,” and he continued, “There are significant differences among dialects, and various local accents and slang abound.” However, in the next edition of the dictionary published in 1886, Hepburn revised the last sentence with “[A]fter the Restoration of Imperial Rule and the transfer of the capital to Tokyo, the Tokyo dialect has gained ascendancy” (13; cited in Lee 2010: 47). During this time when these two editions of the dictionary were published, Japan was experiencing great changes in becoming a modern state, politically, socially and linguistically.

In 1868, after almost 250 years of seclusion, the Meiji government opened the country to the overseas and faced the need to build a modern nation-state (kokka). Pre-modern Japan was composed of over 300 regional clans called han, and in order to synthesize the han into one nation-state, the establishment of a single standard language was considered urgent. The shogunate-domain system (bakuhan taisei) in the Edo period (1603-1868) had prohibited any interactions among han, and this prohibition caused the spoken language to widely diverge into dialects. Additionally, in each dialect, the spoken language had strict distinctions across the four different social classes – samurai, peasants, craftsmen, and merchants, as described by Fukuzawa Yukichi in his Kyūhanjō [Matters of Old Domains] in 1877. At the beginning of the Meiji period, the shogunate came to an end, and the class distinctions as well as the han system were abolished. Holding the idealistic image of an equal society, Meiji linguists, grammarians, language policy makers, novel writers, and politicians believed that the newly established “standard language” should represent the equal “nation-people” (kokumin) and be mutually understood by them. Even though, in reality, there was no such language found anywhere, the
ideology of this imaginary language came to shape what is called *kokugo*, literally meaning “national language.”

In 1894, the year the Sino-Japanese War broke out, Ueda Kazutoshi, a European-trained professor of linguistics at Tokyo Imperial University, delivered a lecture entitled *National Language and the Nation State (Kokugo to Kokka)* in which he introduced the concept of “national language” and emphasized its importance for uniting nations, arguing that “the Japanese language is the spiritual blood of the Japanese people” (1968 [1895]: 110). The following year, Ueda called for the creation of a *hyōjungo*. He defined *hyōjungo* as equivalent to “standard language” in English or *Gemeinsprache* in German, and as that which is “used in a country as a model language” and is, “unlike so-called dialects, understood by most of the people throughout the country” (Ueda 1968 [1895]: 506). Though he did not believe that Japan had ever had such a language, he suggested that with some “refinement,” the Tokyo variety could be the “standard,” since it was the language spoken in the capital of the great empire, and it deserved such eventual recognition (506). Here, it should be noted that Western (i.e. English, French, German) language ideologies were imposed by a European-trained linguist on Japanese culture in the form of a “standard language.”

The language spoken in Tokyo then was a language on a continuum with the “Edo dialect,” which was spoken by lower-class people. Since this native colloquialism was considered to be vulgar, Ueda wanted to disconnect *hyōjungo* from the Edo dialect. Later in *Prospectus for Editing Readers for Normal Elementary Schools (Jinjō Shōgaku Tokuhon Hensan Shuisha)* in 1904, Ueda stated the official definition of *hyōjungo* as “the language spoken in middle-class
society in Tokyo.” In order to urgently realize the standardization, Ueda suggested as the following:

Make the Tokyo language *hyōjungo* as soon as possible and determine it as *kokugo* in its strict sense. And let this be the sole institution that will guide description of the grammar and compilation of common dictionaries. Let it be used in elementary schools throughout the country, and at the same time in all reading, writing, speaking, and listening. ... Thus first establish it as the model language, and then protect it, refine it, and the people will be able to develop it freely (1968 [1900]: 134; translation in Lee 2010: 101-02).

In accordance with Ueda’s proposal, Japanese-language textbooks were prepared for the newly implemented school subject *kokugo* ‘national language’ as part of the effort accompanying the initiation of compulsory education in 1900. According to Lee (2010 [1996]: 107), this change was “a representation of the desire from various facets of society to establish the ideal of *kokugo,*” and “the dissemination of the *kokugo* ideal through elementary schools meant institutionalization of *kokugo* as a normative value to be planted in the people’s consciousness.”

In educational quarters, the enforcement of the Tokyo dialect in the name of *hyōjungo* both fostered an inferiority complex in the minds of the speakers of peripheral dialects and brought about devastating linguistic experiences. While *kokugo* is equivalent to *hyōjungo* linguistically (as declared by Ueda), by calling it *kokugo*, the variety came to embody different metapragmatic ideologies associated with nationalism. Therefore, not speaking *hyōjungo* was considered unpatriotic. The most notorious example is the use of the *hōgen huda* ‘dialect tag’

---

2 Washi (2004: 89) points out that the term ‘middle-class’ then denoted/connoted differently than it does in today’s Japan and referred to rather wealthy upper-class below the nobility and imperial families.

3 Kawaji (2006) notes that the *kokugo* education within Japan developed parallel to the Japanese language education in the colonies and occupied areas outside of Japan.
which was hung around the neck of any student who used local dialectal features at school. The policy of enforcing the Tokyo dialect as hyōjungo and of eradicating peripheral dialects continued until the end of World War II.

1.3.2. Post-Standardization

One hundred years after national efforts to establish the “standard language,” Japanese sociolinguists have commonly reported society’s re-evaluation of regional language varieties. Sanada (2000) claims that language standardization has been completed and that bidialectalism has become more prevalent in contemporary Japan. Kobayashi (2004) proposes the notion of linguistic “accessorization” through regional varieties mainly among Tokyo speakers, where pieces of other regions’ dialectal features are utilized to transform what the Tokyo speakers think of as the “boring standard speech” into a more attractive style. The underlying principle of this notion is the “scarcity value” of regional varieties as a result of the widespread use of the “standard language” ideology (Kobayashi 2004: 106). More recently, Tanaka (2010, 2011) argues that young Tokyo speakers “trinketize” dialectal features by treating them like “toys.” The choice to use the term “toys” instead of “accessories” by Tanaka derives from the most common response to Tanaka’s survey question regarding motivations for using regional features among college students in Tokyo, which was “to create a funny feel [to the utterance].”

In his study of stereotypical images of the Osaka variety (or the Kansai ‘Western’ dialect), Kinsui (2003) argues that while modern fictional work (e.g. literature, mangas, TV shows) typically use the Tokyo variety or the “standard language” for the main characters’ (especially, heroes’ or heroines’) speech, the Osaka/Kansai dialect speakers are depicted as tricksters. Typical attributes of those Osaka/Kansai characters listed by Kinsui (2003: 82-83) are “comedic (joodan-zuki),” “stingy (kechi),” “food-lover (kuishinboo),” “flashy/gaudy (hade-zuki),” “vulgar
“gehin),” “having guts (do-konjoo),” and “yakuza/mafia (yakuza).” However, this convention has been gradually changing since the 2000s, and the Osaka/Kansai variety have begun to represent the voices of main characters in fictional work where their Osaka regionality or stereotypes associated with it are not relevant to the main plots. For example, in the Japanese voice-over of the Disney movie series, *Shrek* (Adamson and Jenson 2001, 2004, 2010), the main character Shrek was performed in the Osaka/Kansai variety by a popular comedian. In 2014, the Hollywood movie *The Secret Life of Walter Mitty* (Stiller 2013) was released in Japan as the very first live-action film where the main character named Walter Mitty, a shy and modest mediocre businessman, played by Ben Stiller, is dubbed in the Osaka/Kansai variety. Whereas the comedic characteristics of Shrek appear to match the stereotypical image of the Osaka variety, the character of Walter Mitty does not link to stereotypes of the Osaka variety. Although the Osaka/Kansai heroic or main characters are still dominantly male in TV shows or movies and rarely female, in the literary world, some award-winning literature (e.g. Nishi 2006, Ogino 2007, Kawakami 2008) utilizes female Osaka/Kansai protagonists’ narratives.

Though the above-mentioned sociolinguistic research on post-standardization treated the Tokyo language variety as a contrast to other regional varieties, it should be noted that the Tokyo variety is one of those regional varieties. As history shows, there has never been such a language variety as a “standard language” in reality. The “standard language” is “an idea in mind rather than a reality – a set of abstract norms to which actual usage may conform to a greater or lesser extent” (Milroy and Milroy 1985: 23), and the “completion of standardization” (Sanada 2000) indicates that the “standard-language” ideology and belief are widely shared in today’s Japanese society.

---

4 The stereotype of the Osaka variety will be discussed at length in Section 1.6.2.
5 Though the Osaka Auntie is a highly established local stereotype in the media, it rarely appears as a main character or a heroine in TV shows or movies. One Osaka participant, Saya, in this dissertation commented that Osaka female characters in TV dramas “spoil the [romantic] mood (muudo buchikowashi).”
1.3.3. Research on Japanese Dialectology

As Milroy (2001: 534) puts it, “dialects cannot be labeled ‘non-standard’ unless a standard variety is first recognized as definitive and central.” This statement was also true in Japan. It was not until the concept of hyōjungo ‘standard language’ was established that the people became aware of the concept of hōgen ‘dialect.’ Once these concepts were established, Meiji intellectuals’ concern was how to modify and refine the Tokyo language into the “standard language” and how to advance “national language” education. In 1902, the Kokugo Chōsakai (National Language Research Council) was founded under the Ministry of Education by Ueda’s proposal. There were four goals of the board: the first two were regarding organizing the writing system, and the other two were about the spoken language. The goals regarding the spoken language were: a) Investigate the phonological structure of the “national language” and b) Investigate dialects for the selection of hyōjungo.

According to Lee (2010 [1996]: 103), the underlying intention of goal (b) was “an investigation into the actual state of dialects with a view toward their extermination,” and goal (a) implied “preparation for correcting the pronunciation of dialect speakers through teaching them hyōjungo.” Under the Kokugo Chōsakai (National Language Research Council) at the beginning of 20th century, dialect divisions were paid great attention. In order to grasp the whole dialectal map of this new nation-state, bundles of isoglosses were sought in order to find dialect boundaries. As is very clear from the goals of Kokugo Chōsakai, extensive research on Japanese dialects arose due to political motives rather than academic ones.

After World War II, research of the Japanese language began to apply highly quantitative survey methods. The first large-scale linguistic survey based on the stratified systematic sampling method was the “Literacy Survey” conducted under the guidance of the Civil
Information and Education (CIE) Section of the General Headquarters (GHQ) of the U.S. Occupation Forces. The Literacy Survey was executed with statisticians at the Institute of Statistical Mathematics in order to prove that the Japanese people had waged a desperate war because low literacy resulting from the complicated system of writing had hindered the flow of accurate information. However, the survey showed instead that the literacy rate of the Japanese was quite high. The experience of conducting the Literacy Survey was very influential to Japanese linguists, and in 1949, when the Japanese government launched the National Language Research Institute (NLRI) to replace Kokugo Chōsakai for the study of the modern Japanese language, those inspired linguists became its first members.

One of the most prominent projects of NLRI in its early years was a series of sociolinguistic investigations of language change, particularly the degree of standardization. Field research was conducted on Hachijo Island in Shirakwa City (NLRI 1950), Tsuruoka City (NLRI 1953), Hokkaido (NLRI 1966), Matsue (NLRI 1971) and several other locations. These research sites were chosen based on various factors such as their unique and prominent dialectal features, similarity of the community size across different sites, and isolation from other neighboring communities. An exact protocol was applied to randomly selected informants in each site. Among these, the study in Tsuruoka, a small northern city, was especially significant since the same surveys were repeated there 20 years later in 1971 (Egawa et al. 1986), and a third time 20 years after that in 1991 (Yoneda 1997) to observe real-time changes and the process of standardization. In the investigation, the informants (total number of 577 in 1951, 457 in 1971 and 205 in 1991) ranging from 15 to 69 years old were asked questions regarding phonetic/phonological, lexical, and grammatical features, as well as educational level, occupation, degree of contact with mass communication, opinions about language, and outside contact. The data obtained was analyzed statistically, using several demographic variables such as gender, age, education level and occupation. In the most recent study, Yoneda (1997)
investigated the phonetic elements⁶ and reports that the age group that showed the most prominent standardization, though not complete, was 25-34 in 1950 and 15-19 in 1971, and that standardization was all but complete by 1991, except for the oldest age groups (45-54 and 55-69). Yoneda further examined lexical intonational patterns and concluded that the standardization of intonational patterns was progressing more gradually than the phonemes.

While gender was included as a demographic factor in the Tsuruoka studies, gender-based variation was not closely examined. The studies simply reported that they failed to reach significant differences between female and male informants in the process of standardization. However, in the early phase of the standardization efforts in Japan, before the Tsuruoka studies, gender-based variation was examined more closely in Japanese research, though no correlation was found between gender and “standard” variants. This is surprising considering that gender-based differences were found in many Western studies of “standard” varieties. Before moving onto those Japanese studies that focus on gender, the following section briefly reviews Western research on gender and “standard” varieties.

1.4. “Standard” Varieties and Gender in Western Communities

A number of studies on Western communities have repeatedly shown that women make greater use of more “standard”⁷ variants than men for stable variables (Labov 2001), that is, those variables that are well established in a community and are not undergoing change (Chambers 2003: 121). One of the earliest studies to note the association between women’s speech and “standard” variants is Fischer (1958). In this study, Fischer compared male children’s and

---

⁶ The phonetic elements investigated include labialization, palatalization, voicing, nasalization, centralization of consonants, and merger of the front vowels.

⁷ While the notions of “standard” (or “prestige”) and “nonstandard” variants are taken it for granted in many sociolinguistic studies, I here use quotation marks to indicate that the those terms are themselves ideologically loaded as Coupland (2010: 60) emphasizes.
female children’s use of the –ing variable (e.g. walking vs. walkin’) in a semi-rural New England village and found that female children consistently used more of the “standard/prestige” form [ðŋ] than male children. Since this finding, various proposals have been suggested to account for the different patterns between women’s and men’s use of “standard” speech, a number of which maintain that the usage patterns are due to women’s higher concern for “prestige” due to their relative powerlessness in many communities.

The notion that women’s higher concern for “prestige” in the use of “standard” speech is clearly presented in Labov’s (1966) New York City study, Wolfram’s (1969) inner-city Detroit study, and Trudgill’s (1974) Norwich study. Labov’s study found that while women used fewer stigmatized forms (such as r-less pronunciation) in careful speech, in shifting between casual and careful styles, women showed sharper shifts from stigmatized forms to “prestige” forms (such as r-ful pronunciation) than men did. Labov (1966) interpreted the women’s sharp style shifting as the result of “hypercorrection” (312) due to their “linguistic insecurity” (335). Labov (1972: 243) returned to this point later and revised it, saying that “women are more sensitive than men to overt sociolinguistic values.” Wolfram (1969) examined four phonological variables and four grammatical variables used by 12 (6 males, 6 females) African Americans in each of four social classes (upper middle, lower middle, upper working, and lower working). He found gender correlations for all eight variables across the social classes and concluded that “females show a greater sensitivity to socially evaluative linguistic features than do males” (76).

In his Norwich study, Trudgill (1974) conducted self-assessment tests in which the informants were asked to listen to two different pronunciations (“standard” pronunciation and “nonstandard” pronunciation) of the same word and indicate which of the pronunciations more closely resembled the way in which they normally produced the word. The results of the test indicated that female Norwich informants were much more prone to over-report their use of “standard” pronunciations while male informants were much more likely to under-report, which
led him argue that women and men respond to opposed sets of norms. While women respond to overt, “standard-language” “prestige” norms, men respond to covert, “vernacular” “prestige” norms. Trudgill speculated that women’s “prestige”-consciousness was a result of their subordinate position in the community compared to men’s. Since women have fewer opportunities to secure their status through their skills, occupational success, and other outlets, they find it necessary to use symbolic means, such as language, to enhance their position.

Gal’s (1978) and Nichols’ (1983) studies present cases where women’s “standard” speech is not merely a symbolic means for their upward social mobility. Gal (1978) reported that young women of the Hungarian-speaking peasant class in a village in Austria used German, the “prestige” language, significantly more than young men. Gal explains that the social oppression that the young women are subject to in the community gives them less motivation to maintain loyalty to Hungarian while at the same time they have better chances to attain easier lives by marrying into the German-speaking urban middle class. Nichols (1983) observed that rural African American young women’s speech in a working-class island community in coastal South Carolina exhibited more “standard” features than younger men’s speech due to the occupational choices available for women. In the community, the average salary for black women is less than half that for black men, but white-collar jobs typically available for women that require knowledge and use of “standard” English such as sales, nursing, and teaching pay almost as much as some blue-collar physical labor for men. Therefore, the “standard” language used by women in this community indicates women’s incentive to train for the higher-paying jobs.

Though women’s “prestige”-consciousness offers a reasonable explanation for female-male differences in the use of “standard” variants generally, it does not provide sufficient explanation for intra-gender differences. The above-mentioned Nichols’ (1983) study of African American residents in coastal South Carolina showed variation among women by including different age groups and two sub-communities, island residents and mainland residents, in
addition to the female-male differences. While younger island women used more “standard”
variants than younger island men as is noted earlier, older island women used approximately
the same degree of “nonstandard” variants as older island men. Furthermore, mainland older
women exhibited the highest percentage of “nonstandard” forms among all the groups in the
study, while using more than twice as many “nonstandard” variants as older mainland men. To
understand these women’s “atypical” patterns, Schilling (2011) notes the importance of
additionally considering differences in the localized gender roles and economic circumstances in
the particular communities. Nichols’ ethnographic investigation showed that the inter-gender
and intra-gender differences were results of very different educational and occupational
experiences available to different gender and age groups. For example, the older women’s
frequent use of localized “nonstandard” features is because of their lack of schooling as well as
lack of exposure to the “standard” features through jobs such as domestic and farm labor,
whereas the younger women’s increasing use of “standard” features is a result of newly available
women’s jobs that require “standard” English. This shift shows that individual linguistic
patterns change as the gender roles and economic conditions change, which, in turn, suggests
that those conditions resulted in different speech patterns among these women.

Intragender variation is also observed in adolescence, an age group where occupational
choices do not normally provide a strong motivation for people’s linguistic practices in the
United States. In her ethnographic study of Detroit suburban high school students, Eckert
(2000) showed that the use of new urban sound changes and a “nonstandard” morphosyntactic
feature, negative concord, distinguishes the two main opposed social categories in this high
school: burnouts and jocks. The burnouts reject school and orient themselves to the local and
urban area, while the jocks actively participate school activities, avoiding the local urban area.
The opposition between burnouts and jocks is manifested by burnouts’ use and jocks’ avoidance
of urban sound changes and negative concord. When considering gender categories, it is clear
that the burnout girls are the most vernacular speakers, whereas the jock girls are the most conservative and “standard” speakers, with the jock and burnout boys falling between them. In particular, the most vernacular speakers in the data are a group of girls known to be the “wildest” burnouts. This study reveals that the use of “nonstandard” variants can be as significant as the use of “standard” variants in construction of female gender identities.

This section has briefly reviewed some foundational work on “standard” varieties and gender in Western communities, but it should be noted that there are many other studies from more current viewpoints; I address these below in Section 1.5.2. Unlike these Western studies of gender and “standard” varieties, research on Japanese dialects like the Tsuruoka studies have not extensively examined male-female differences in the use of “standard” variants. However, in the early phase of the standardization efforts in Japan, before the Tsuruoka studies, differences between women’s and men’s speech were one of the governments’ major concerns. While hyōjungo was meant to represent the modern male voice (Okano 1964 [1902]: 510), around the same time, women’s speech came to be represented by the language variety called Japanese Women’s Language (JWL). Just like hyōjungo, JWL was a target of national attention because of complex historical and political forces revolving around modern nation building. The national project of standardization during the modernization movement had a different significance for women than for men. We now turn our focus onto the other relevant field of this dissertation, Japanese Women’s Language.

1.5. Relevant Field II: Japanese Women’s Language

Japanese Women’s Language (JWL) is a language variety that embodies a specific kind of linguistic ideology. Linguistic ideology is defined as “sets of beliefs about language articulated by users as a rationalization or justification of perceived language structure and use” (Silverstein
1979: 193) or “the cultural system of ideas about social and linguistic relationships, together with their loading of moral and political interests” (Irvine 1989: 255). While these definitions imply that the “beliefs” or “ideas” are inaccurate or distorting in some way, Agha (2007: 396) argues that notions of distortion should be more explicitly incorporated in definitions of ideology. This particular Japanese ideology offers sets of “feminine” linguistic features and at the same time “masculine” linguistic features as well as distorting linguistic beliefs about forms and functions of language used particularly by women, such as “how Japanese women usually speak” or “how Japanese women should speak.” The inaccuracy and distortion of this ideology is evident in the recent studies of Japanese women’s linguistic practices reviewed in Section 1.5.2 below.

1.5.1. Socio-Historical Context of Japanese Women’s Language

One of the common myths about Japanese Women’s Language (JWL) is that it is a “cultural heritage.” M. Inoue (2002, 2004, 2006) and Nakamura (2008) discuss how JWL was created by male intellectuals during the Meiji (1868-1912) modernization period in the late 19th century, and imposed as a hegemonic ideological construct on women as the ryosai kembo ‘good wife, wise mother’ in order to “nationalize” women (Ueno 1998). The agenda behind the modern “ideal” women or ryosai kembo sharply differed from the pre-modern doctrine that simply taught women to be obedient to their husbands. Under the new national agenda, “household management (kasei)” and education of children, which had both previously been considered men’s work, became the most important facets of women’s work. In order to perform the role of educating their children, women were encouraged to study and enrich their knowledge, and it
was 1899 when institutions of “higher education” for women older than 12 years old were founded.8

Sawada Masako, who dedicated her life to ryosai kembo education and founded a women’s educational institution, published a few conduct and etiquette manuals for women (e.g. Joshi no Honbun [Women’s Primary Duties] in 1894, Joshi Kyooiku Yoogen [Essential Points of Women’s Education] in 1897). In those manuals she repeatedly advocated the importance of women’s virtue (jo-toku), intelligence (chiryoo), and elegance (hinkaku) (e.g. Sawada 1894: 129). Additionally, other writers of conduct and etiquette manuals, mainly male writers, emphasized the “ideal nature” of women as being “motherly,” “gentle,” “affectionate,” “modest,” and “self-sacrificing” (Kurasumi 2008).

During this Meiji (1868-1912) modernization period, JWL was established and propagated to embody this “ideal” ryosai kembo figure as a critical part of the process of modern nation building. Among the features of the newly created JWL were feminine self-referential forms, beautifying prefixes o- and go- (e.g. o-sushi ‘sushi,’ go-han ‘rice’), honorific expressions, as well as the use of new word-final particles to be used by women in place of traditional particles used by people of both genders. M. Inoue (2004) investigates final particles used by female characters in two popular fictional works (Sanba (1952 [1813]) and Natsume (1985 [1909])) written before and after the Meiji Restoration (1867) and reports that while the particles in the pre-Meiji work are not gendered, particles in the post-Meiji work display the systematized and standardized gender marking of the modern nation.

Much of JWL has its roots in so-called “schoolgirl speech.” M. Inoue (2006) highlights the process by which schoolgirls’ speech (jogakusei-kotoba) was elevated to “women’s language” during the Meiji modernization period. Schoolgirls referred to “girls and young women of the

---

8 These are the original forms of women’s colleges today.
elite classes who attended the women’s secondary schools that had been instituted as part of the early Meiji modernization project” (M. Inoue 2006: 38) and represented a new social category of female with their “modern” image. However, initially, when their speech was first publicly heard, there was a major criticism by Meiji male intellectuals about how vulgar schoolgirls’ speech sounded. Schoolgirls were reported to use a set of distinctive final particles, such as -teyo, -dawa, and -noyo, whose similarity to the final particles used by yuujo (literally meaning ‘women who play’) who in pre-Meiji eras worked alongside geishas (literally meaning ‘women with artistic skills’) and provided sexual services was often noted. Yuujo was a euphemistic title given to prostitutes in pre-Meiji eras, but the profession of geisha around this time period was inconsistently defined, with some claiming that geishas were simply performing artists and others implying that they offered sexual services like yuujo. In 1909, a scholar of Buddhism and Japanese culture, Arthur Lloyd, described geishas as “dancing-girls” and noted that “[t]he line of demarcation between the geisha and the [yuujo] is sometimes very indistinctly traced.” Some Meiji male intellectuals, with a hint of sarcasm, made a point that the schoolgirls learned use of the features from their mothers, ex-geishas who were married to men of status during the time of social upheaval.

The unique final particles and other features used by schoolgirls became the essential linguistic features that define idealized women’s language through their use in modern literature. In the literary world, Meiji writers were developing their modern narrative prose by translating Western realist novels. Those writers needed to find the authentic Japanese speech of modern women to represent the translated voice of Western white female characters. As an easy solution,

---

9 M. Inoue (2006: 42-43) argues that schoolgirls were different from other “educated women” from the pre-modern eras in that they were symbolic of the modern and national women. She notes, “[Schoolgirls] were to be sighted in public space, particularly in modern space, as iconic figures essential to the new urban landscape, including parks, department stores, museums, zoos, train stations, and downtown streets.”

10 However, Lloyd added that there are some geishas who possess manner, intelligence, and class that come near “Aspacia, the chosen companion of Pericles,” one of the most influential Athenian political figures (cited in Ueda 2013: 11). Lloyd’s remarks suggest that around this time, there are mixed views of admiration and disdain toward the geisha.
they turned to the female exclusive variety they had overheard on the street – schoolgirls’ speech. While this realist movement was an aesthetic pursuit in literary circles, it led to tangible outcomes, in that readers believed that the fictional narratives faithfully recorded the real speech of modern women. Readers enjoyed the new narratives while consuming the “realistic” image of women of the urban middle and upper classes (M. Inoue 2006: 101).

Once the newly made women’s language along with the image of modern women was established in people’s minds through the literature, its ideology came to be more explicitly propagated through conduct and etiquette manuals. Propagation of JWL among women was, in fact, a critical part of the establishment of “national language” and language standardization among men. Nakamura (2008) emphasizes that the implicit masculinization of “national language” was accomplished by its relationship to marked, exceptional, and marginal language ideologies associated with women. In contrast to the male-spoken standard language discussed in academic, political, and literary discourse as a topic crucial to national unity, women’s speech was mainly discussed in the discourse of conduct and etiquette manuals through print media (M. Inoue 2006, Nakamura 2008).

While JWL was imposed upon all Japanese women, the propagation of JWL appeared to be less successful than the standardization. As noted earlier, Japanese dialectological studies of language standardization statuses like the Tsuruoka studies did not closely examine gender issues, but gender-based differences in speech were, in fact, one of the government’s concerns at the outset of the project. Central to the agenda of the very first nationwide dialect survey conducted by the Kokugo Chōsakai (National Language Research Council) was investigation into differences in the use of speech forms according to gender as well as position in the local social-class hierarchy (M. Inoue 2006: 103). While the results showed some gender-based and class-based differences in the city of Tokyo, such sociolinguistic differentiation was not reported in outlying areas, which was unexpected by the elite Tokyo-based members of the Kokugo
Chōsakai. Six years later, the council planned a second survey with a design in which the respondents were directed to pay attention to the specific sociolinguistic factors including gender, although the results were unknown since they were never published as an official document. In the following section, I will provide an overview of literature on JWL and highlight how JWL failed to be incorporated into Japanese women’s daily speech.

1.5.2. Research on Japanese Women’s Language

The academic study of JWL emerged in the late 1970s when female Japanese linguists were inspired by Western research on language and gender. Following the landmark studies by Lakoff (1975) and Spender (1980), in the first two decades of its history, research on JWL was primarily concerned with *ideological* differences between “women’s language” and “men’s language” and treated gender as an essential attribute. Scholars attempted to connect specific linguistic forms or styles directly to speaker’s gender categories based on the ideology, theorizing women’s speech as either reflective of dominance by men or of cultural differences from men. The linguistic features commonly examined as stereotypical JWL features in those studies include the use of polite expressions (Adachi 2002, Ide et al. 1986, Farnsley 1995, Okamoto 1999), feminine vs. masculine self-referential and address terminologies (Ide 1979, Shibamoto 1985, Shbatani 1990, Kanemaru 1997, Uchida 1997), feminine vs. masculine phrasal-ending particles (McGloin 1997, Reynolds 1985, Kawguchi 1987, Ide 1979), interjections/exclamatory expressions (Jordan 1974, Shibamoto 1985) and voice pitch (Ohara 1992, 1997, Van Bezooijen 1995, Yuasa 2011). Though the research was greatly influenced by work written in English, some of these studies on JWL were less concerned with feminism than was Western research. Ide (1993), who claimed to aim for a “scientific study” not motivated by feminism, argued that JWL, rather than being a tentative, powerless, and trivial language, as feminists tended to characterize
it, displays dignity and that Japanese women have more actual power in domestic matters than men (Ide 1979, 1993). Though this view received strong opposition by some feminist scholars (e.g. Reynolds-Akiba 1993), the approaches of both feminists and non-feminists drew on and consolidated the clear-cut dichotomy of “women’s language” and “men’s language”.

In the 1990s, a new theoretical framework emerged in Western studies of language and gender, based on the development of social theories and discourse-based methodologies. This new approach views gender as a social and interactional construct rather than a given attribute. Eckert and McConnell-Ginet (1992) viewed the interaction of gender and language as jointly constructed in the everyday social practices of particular local communities, called communities of practice. Gal (1995: 180) observed, “the study of language and gender is significantly enhanced by simultaneous attention to everyday practices on the one hand, and on the other to the ideological understandings about women, men, and language that frame these practices and render them interpretable in particular social contexts, historical periods, and social institutions.” Applying this new theoretical framework, JWL researchers began to pay more attention to linguistic practices in addition to the language ideology, and many of more recent studies (e.g. Okamoto and Sato 1992, Okamoto 1995, Miyazaki 2002, 2004, Matsumoto 2002, 2004, Millar 2004, Lee and Yonezawa 2008) discovered that the ideology is not necessarily reflected in the practices. For example, Okamoto (1996) examines the use of the beautification honorific prefix and sentence-final ending forms in dyadic conversations between women in two age groups and found that younger women use both honorific prefixes -o (e.g. o-sushi vs. sushi) and feminine sentence-ending forms (e.g. -wa) less frequently than older women. Matsumoto’s (2004) empirical study reports that even middle-aged women in her study do not use most stereotypical feminine sentence-final endings very often. Miyazaki (2002, 2004) highlights that some junior high school girls in her study use masculine self-referential terms (e.g. boku, ore) instead of feminine forms (e.g. watashi, atashi). In her investigation of gyaru in Tokyo, a group
of young girls who are identified with a specific fashion style characterized by heavily dyed hair, decorative nails, and dramatic make-up, Miller (2004a) argues that these shockingly rebellious high school girls resist using conventional features of JWL.

While the findings of these JWL studies were groundbreaking in the area, these studies overlooked a significant point regarding gender construction: construction of gender is more dynamic and multifaceted than just using (or not using) masculine or feminine linguistic features. Gender construction has every bit as much to do with constructing orthogonal dimensions of identity (Podesva and Kajino 2014: 111) such as region and age. In order to highlight this point, my study focuses on the phonetic features that have not previously been documented as JWL features. I will demonstrate how Japanese women utilize resources other than those ideologically marked as more or less feminine or masculine to construct their gender identities.

1.5.3. Japanese Women’s Language and Region

To date, the majority of previous studies of JWL, including the ones reviewed earlier, discuss only the variety spoken in the Tokyo metropolitan region, and there is little research that concerns regional varieties. Examining the discourse of three Japanese women in a farming community in the Ibaraki region, Sunaoshi (2004) has shown that JWL is not part of the speech repertoire of those women, while those women are aware of the ways in which JWL indexes middle-class femininity. Sturtz-Sreetharan’s (2008) empirical study has examined morphological features in the speech of nine women in the Greater Osaka region, mainly from Kobe to Osaka, and argued that the participants used Tokyo forms as a means of distancing themselves from the negative connotations of the stereotypical tough and loud Osaka women but simultaneously used regional forms to position themselves clearly as regional women and
not users of JWL (the regional stereotype will be discussed at length later in this chapter). For the preliminary study (Kajino to appear) for the current dissertation, I quantitatively analyzed phrase-final particles using sociolinguistic interview data recorded from 1977-83 (National Institute for Japanese Language and Linguistics (NINJAL) 2002a, 2002b, 2002c) with 13 speakers from Tokyo, Kyoto, and Osaka (one female and one male from Tokyo; two females and three males from Kyoto; and two females and four males from Osaka) ranging in age from 49 to 79 years old. The results revealed that the JWL forms were disfavored by the Kyoto and Osaka women while they were favored by the female Tokyo speaker.

The lack of JWL research on non-Tokyo varieties is partly related to the methodological challenges researchers confront when they attempt to analyze linguistic practices of women from other regions. The above-mentioned common approach taken by JWL studies (i.e. attempting to account for linguistic practices based on the conventional feminine-masculine dichotomy) becomes problematic in analyses of non-Tokyo varieties, since the feminine-masculine dichotomy of JWL is not even applicable in linguistic practices of non-Tokyo women. As is shown by Sunaoshi (2004), Sturtz-Sreetharan (2008), and Kajino (to appear) as well as the historical Kokugo Chōsakai (National Language Research Council) dialect surveys, JWL has never been implemented into non-Tokyo women’s speech.

In the next section, I will outline stereotypical images of the Tokyo, Kyoto, and Osaka language varieties as well as JWL described by the participants of the present dissertation. In the interviews, the participants were asked to elicit their views about these varieties in the form of free discussions. The metalinguistic commentaries in their responses clearly portray the distinctive gendered stereotypes of the three varieties that present an intersection of gender and region.
1.6. Participants’ Attitudes toward Regional Varieties and JWL

During the interview sessions in this study, I asked each participant the same set of questions. In this section, I will focus on the participants’ responses to the following questions in order to examine their attitudes toward the Tokyo, Kyoto, and Osaka varieties, JWL, and their own speech. A complete questionnaire listing all the questions asked in the interviews is included as Appendix II at the end of this dissertation, and more detailed information about the interview setting will be provided of Chapter 2.

- Do you think you speak the Kyoto/Osaka/Tokyo variety?
- Are you proud of speaking your own variety? Why/why not?
- What kind of images do you have toward the Kyoto/Osaka/Tokyo varieties?
- Among the three varieties, which do you think is the most feminine/masculine?
- Do you think you use JWL? If not, why don’t you use it? If yes, do you use it all the time? In which situations do you use it? Why do you use it?
- What kind of images do you have toward JWL or women who use JWL?

While language attitudes toward regional dialects and JWL provide critical background information in this study, there is no previous work that discusses those issues thoroughly. Therefore, I present my study of language attitudes toward women’s and regional varieties in Japan here, in a section of the introductory chapter.

1.6.1. Feminine and Masculine Ratings

During the interviews with the 46 participants in this dissertation, I asked questions to elicit their images of the Tokyo, Kyoto, and Osaka varieties. Asking these questions was motivated by a previous large-scale survey study conducted by F. Inoue (1988). Inoue discussed the variety of
Nagoya in Central Japan in addition to the varieties of Tokyo, Kyoto, and Osaka, but the discussion of the Nagoya variety is put aside here since it is not relevant to my study. In F. Inoue’s study, seven native-born informants were selected from each of seven age groups from teens to 70’s in 71 localities (total N=3,479) along a railway line stretching for about 400 km between Tokyo and Osaka. Gender of the informants is not reported in the study. Using a questionnaire, the informants were asked to rank the language varieties of Tokyo, Kyoto, and Osaka based on the levels of standardness (similarity to the “standard language”) and elegance. These two qualities were chosen to investigate participants’ intellectual and emotional attachments toward the local language varieties. The informants were divided into three groups according to the area, Eastern, Central, and Western, and average scores of the rankings in each area were calculated by assigning score 1 for the highest ranking and score 4 (because the ranking included the Nagoya variety) for the lowest ranking for each informant’s response. For the levels of standardness, informants from the Eastern, Central, and Western Japan agreed that the Tokyo dialect is the most “standard” variety, but the ranking of the Kyoto and Osaka varieties varies depending upon the informants’ home areas. Informants from Eastern Japan ranked Kyoto and Osaka as equally “nonstandard,” but informants from Central Japan responded that Kyoto is more “nonstandard” than Osaka. Conversely, informants from Western Japan perceived Osaka was more “nonstandard” than Kyoto. The resulting ranking regarding the levels of elegance was more consistent. Regardless of the locations of the investigation, the informants agreed that the Kyoto variety was most “elegant”, Tokyo was the second, and the Osaka variety was the least “elegant” variety. It is noteworthy that though Kyoto and Osaka are categorized under the same dialect region, the dialectal images of the two varieties in terms of “elegance” diverge widely. Considering elegance is a crucial quality of traditional Japanese womanhood (see discussion of ryosai kembo in Section 1.5.1), this pattern suggests that Kyoto and Osaka carry very distinctive gendered images.
The participants in my present study further confirm the distinctive, possibly gendered, images of the three varieties. As discussed in more detail in Chapter 2, the participants are 46 female college students from the designated regions, including 15 from Tokyo, 15 from Kyoto, and 16 from Osaka. Building on F. Inoue’s (1988) findings, to examine more specific gendered images, I asked the participants to rank the three varieties based on levels of femininity and masculinity with the questions, “Among the three varieties, which do you think is the most feminine (onna rashii ‘feminine’ and onnnappoi ‘effeminate’) /masculine (otoko rashii ‘masculine’ and otokoppoi ‘manish/manly’)?” By using the adjectives onnnappoi ‘effeminate’ and otokoppoi ‘manish/manly’ in addition to onna rashii ‘feminine’ and otoko rashii ‘masculine,’ I intended to evoke both female and male figures for the both feminine and masculine ratings. These ratings were collected from 14 Tokyo participants excluding one who said she did not know how to answer, all the 15 Kyoto participants, and all the 16 Osaka participants. Average scores of the ratings were calculated based on each participant’s response where the most feminine/masculine rating was assigned score 3 and the least feminine/masculine rating was assigned score 1. Note that all of these participants in my study are female while F. Inoue’s study does not specify speakers’ gender.
Figure 1.2 presents the average scores of ratings in femininity. The X-axis shows the home regions of the respondents, and the Y-axis shows the average score of the ratings, with 3 as the most feminine and 1 as the least feminine. It is clear from the figure that across the three regions, Kyoto is consistently ranked as the most feminine, and Osaka is ranked as the least feminine. This finding is compatible with the ranks of “elegance” presented in F. Inoue (1988). Furthermore, it is interesting to note that the participants in each region rate their own varieties as slightly less feminine compared to the ratings by those in other regions.
Examination of the socio-historical context reveals that it is not a very recent trend to consider the Kyoto variety as elegant or feminine. During the Meiji modernization period, Otsuki (1905: 17) compared Tokyo and Kyoto and claimed that the Tokyo variety should be standard because “the language of Kyoto, though good for women, can sound weak coming from men” and “has exaggerated intonation. [The language of Kyoto] is not suitable for giving orders to soldiers, criminal examinations by a judge, or telephone conversation.”

Whereas the participants in each area tend to rate their own varieties as less feminine compared to the ratings by those in other regions, the participants also tend to rate their own varieties as more masculine than those of the other areas. Figure 1.3 presents the average scores of ratings in masculinity. While Osaka is consistently ranked as the most masculine across the three regions, the rest of the rankings vary between Kyoto and the other two regions. The Tokyo and Osaka participants rank the Tokyo variety as the second most masculine and the Kyoto variety as least masculine, but the participants from Kyoto reverse the rank by rating their own variety as more masculine than Tokyo.
Figure 1.3. Masculinity ratings among the three varieties

These participants’ tendencies to rate their own varieties as less feminine and more masculine may indicate the gap between their self-images of their speech and the stereotypical/traditional feminine image. The tendency is most clearly observed in the masculinity rating for Kyoto, which is considered as the most feminine variety. Unfortunately, it is unclear from the responses if the participants were describing Kyoto men’s language use or Kyoto women’s language use for the ratings, but considering the Kyoto participants also rated their own variety less feminine for the feminine ranking, they were probably rating both Kyoto men’s and women’s language use here. The Kyoto participants are aware of the feminine image of their variety, but they think that
their own speech is less feminine than the image, which will be evident from the metalinguistic commentaries provided in responses to other questions. We now turn our attention to those metalinguistic commentaries provided by the participants.

1.6.2. Metalinguistic Commentaries and Local Stereotypes

Metalinguistic commentaries obtained through interviews with the participants of this study portray the distinctive stereotypical images of the varieties spoken in Tokyo, Kyoto, and Osaka. During the interviews, I asked the participants to elaborate on the images they have toward three varieties by asking “What kind of images do you have toward the Kyoto, Osaka, and Tokyo varieties?” The length of responses varied between the participants from simply one adjective or noun to a few sentences using several adjectives/nouns. All the adjectives and nouns used to describe the language varieties are presented in word clouds below. The word clouds were created with Wordle (http://www.wordle.net/) based on the frequency of the word occurrences in the commentaries.

Figure 1.4 shows the word cloud of “the Tokyo variety (Tokyo-ben)” based on the metalinguistic commentaries. The Tokyo variety was most commonly (by 13 participants) described as being “cold (tsumetai),” which is linked to the image of being “distant (kyori ga aru, kabe ga aru).” The terms “normal (futsuu)” and “effeminate (onnappoi)” are both used by 7 participants, but while 6 out of 7 who used the term “normal” are Tokyo participants, all the participants who used the term “effeminate” were non-Tokyo participants. Additionally, for the Tokyo participants, their own variety is “normal,” but “boring (tsumaranai)” and has “no charm (aiso ga nai, kosei ga nai),” whereas the Kyoto and Osaka participants commonly “disfavor (iyana, kirainai)” the Tokyo variety, particularly male Tokyo speech since it sounds “effeminate” and “wimpy (taylorinai, nayonayo shiteru).” On the other hand the Tokyo female speech sounds
“cute (kawaii)” to the non-Tokyo participants. This effeminate image of the Tokyo variety suggests that there may be some link between “standardness” and femininity in Japanese although the association between two has not been reported in previous literature on Japanese. Furthermore, the Tokyo variety is the language used in “TV” shows, such as “dramas,” and “animes (‘animated cartoons’),” and to the Kyoto and Osaka speakers sounds “fake (wazato-rashii).” Here, it might seem odd that qualities such as “cold,” “distant,” “effeminate” and “cute” could co-exist in the image of the Tokyo variety, since cuteness has often been discussed in relation to intimacy in previous literature (e.g. Kinsella 1995, Miller 2004, Moon 2010, Kajino 2011). However, this seeming mismatch shows that the notion of cuteness changed in Tokyo and Japanese society. “Coldness” is an essential element of a new type of cuteness, which I will discuss at length as “urban cuteness” in Chapter 5.

Figure 1.4. Word cloud for the image of the Tokyo variety
While the commentaries on the Tokyo variety differ between respondents in Tokyo and the other regions, the images of the Kyoto and Osaka varieties are more consistent across the regions. Being “classy (joohinna)” (with 10 responses) is the most prominent image of “the
Kyoto variety (‘Kyoto-ben’),” with this quality in turn being associated with the words like “hannari,” “geisha”, and “traditional Japanese.” Hannari is a locally specific adjective/adverb used mainly in Kyoto and represents qualities such as “soft (yawarakai, yanwari),” “calm (ochitsuiteru)”, and “quiet (shizuka)” beauty. This particular term was used only by the Kyoto and Osaka participants in the responses. As is argued by Foreman (2005), the widely accepted image of geisha in today’s Japanese society does not resemble what is known as the “geisha-girls” (i.e. high-class prostitutes)\(^\text{11}\) in the West. As we discussed earlier, in the past, the profession of the geisha was vaguely defined, but the notion of geisha in today’s Japan is clearly distinct from the various other women who, previously in history, provided sexual services, including yuujo and shoogi (post-Meiji title for yuujo; literally meaning ‘prostitutes’) (Foreman 2005: 34-35).\(^\text{12}\) Considering that their younger counterpart (maiko) is an apprentice geisha, the term “geisha” also connotes some maturity. As is clear from all the positive descriptive terms, the Kyoto variety is “favored (sukina)” by many participants, but some negative descriptions include being “two-faced” and “dishonest.”

Finally, “the Osaka variety (‘Osaka-ben’)” has the most consistent images. Some negative images are being “harsh (kitsui),” “pushy (gangan kuru),” “tough (tsuyoi),” “straight-forward (hakkiri iu, sutoreeto),” and “bad mouthed (kuchi ga warui),” and other positive images include “friendly (shitashimi yasui, nakaii kanji),” “cheerful (akurai),” and “funny (omoshiroi).” The “funny” image is derived from the fact that Osaka has been the home of one of the largest entertainment agencies in Japan since the early 20th century. This agency is best known for its popular comedians, who commonly speak the Osaka variety (and western dialects) on their

\(^{11}\) Ueda (2013:11) shows how geishas historically have been considered prostitutes in Western culture by quoting a travel book written by a French journalist, Dekobra (1936): “[E]veryone who has visited Japan expects to be asked, on his return, the inevitable question, a question which always brings a smile of the questioned: ‘And the geishas?’.” Such images persist today, as exemplified, for example, in the popular Hollywood movie Memoirs of a Geisha (Marshall 2005) also depicts the geisha house where the main character grows up as if it were a brothel.

\(^{12}\) Buddhist scholar Arthur Lloyd noted that the distinction between the geisha and the prostitute could be vague in 1909 in Japan, but approximately 50 years later, a French-Japanese author, Yamata (1956) stated that “[a geisha] is not a prostitute” and translated the word ‘geisha’ as “artiste.”
television shows, so the word “comedian” was frequently used to describe the variety. “Osaka Auntie (Osaka no Obachan)” is a highly enregistered local character type (Agha 2007) which is often presented in comedic portrayals in the mass media\(^\text{13}\) (Kajino 2014) and popular books (Maegaki 2005, 2014). Figure 1.7 shows Osaka Auntie illustrations on the book cover of Osaka no Obachan Korekushon [Osaka Auntie Collections] (Maegaki 2014). The book comedically depicts a “typical” Osaka Auntie figure in terms of her flashy fashion styles, frank and friendly behaviors, and daily practices.

![Osaka Auntie illustration](image)

**Figure 1.7.** Osaka Auntie illustrations on the book cover of Osaka no Obachan Korekushon [Osaka Auntie Collections] (Maegaki 2014)

Again, it should be noted here that the varieties of Kyoto and Osaka are categorized under the same dialect region, but the dialectal images of the two varieties diverge widely. Whereas the Kyoto figure portrayed by the commentaries is a classy and calm geisha, the

\(^{13}\) Tanaka (2011: 68-91) discusses the role of the mass media in the construction and circulation of stereotypes associated with local language varieties by examining contemporary literature and mangas (‘comic books’) as examples.
stereotypical figure in Osaka is a harsh, bad-mouthed, and funny auntie. Compared to the Kyoto and Osaka varieties, the stereotypical female figure of the Tokyo variety is vague from the participants’ responses. This is because enregisterment of the Tokyo variety as the “standard language” deletes connotations of “localness.” It has been documented that the absence of regional affiliation is a component of “standardness” (Ladegaard 2001). Another reason why the Tokyo female figure is not very clear is that the images portrayed by the participants vary across the regions. The Tokyo participants depicted a “cold and boring” figure without any charm, while the Kyoto and Osaka participants drew a “distant but cute” figure just like the female characters that appear on TV. One of the Osaka participants (Miya) distinguished the three regional female figures by stating the Tokyo female figure shows “a trendy (imadokippoi) femininity,” while the Kyoto female figure shows “a traditional (kofuuna) femininity,” and the Osaka female figure shows “a unique and eccentric (koseitekina) femininity.” This dissertation demonstrates that Miya’s characterizations seem to be quite apt. In Chapter 5 in particular, we will see how participants use linguistic and extralinguistic style to construct images associated with the Tokyo “urban cuteness” that emerges from the metalinguistic commentaries obtained from the non-Tokyo participants. However, the boring standard image of their own variety is more evident among the Tokyo participants, and, as we will see, they have to negotiate this stereotype alongside the widespread image of “urban cuteness” in the construction of their gender identities.

1.6.3. Self-Evaluations of Regional Speech

While the metalinguistic commentaries on the Tokyo, Kyoto, and Osaka varieties present established local images, the participants’ evaluations of their own speech do not necessarily match with the local stereotypes. The mismatch is particularly noticeable in Kyoto. To the
question “Do you think you speak the Tokyo/Kyoto/Osaka variety?” responses varied across the regions. Though it has been shown in previous literature (Labov 1966, Trudgill 1974) that people’s self-evaluations of their own speech do not always accurately capture their linguistic practices, this question is helpful to observe participants’ views on their own speech and the local stereotypes. To the question, 12 out of the 15 Tokyo participants answered that they think that the variety they speak is “hyōjungo (‘the standard language’)” and not “the Tokyo dialect (‘Tokyo-ben’).” These responses indicate that the Tokyo participants do not consider the Tokyo variety as a regional dialect. Two Tokyo participants did not know the answers to the question because they “haven’t thought of it,” and one said that her speech displayed mixed dialectal features due to the influences of other people’s regional speech. In Kyoto, 12 out of the 15 participants said that they did not think they spoke “the Kyoto variety (‘Kyoto-ben’),” but instead, they spoke “the Kansai (‘Western Japanese’) dialect (‘Kansai-ben’).” One of the two who responded that they “probably” spoke the Kyoto variety realized that when other people had told them that they did speak the Kyoto variety. The last Kyoto participant rejected answering the question since she did not know what the actual speech of “the Kyoto variety” sounded like. In contrast, to the same question in Osaka, all of the 16 participants agreed that they spoke “the Osaka dialect (‘Osaka-ben’)” without any hesitation. Some even emphasized the positive answers as “I very much think so (meccha omou)” or “Yes, definitely (hai, gattsuri).”

The fact that most of the Kyoto participants rejected the idea of speaking the Kyoto variety indicates their perceived mismatch between the stereotypes attached to the Kyoto variety and their self-images of own speech. This indicates that while the Kyoto variety carries the classy, soft, and pretty geisha-like image, the participants think that their own speech does not match up to the traditional feminine image. Rather, by labeling the variety they speak as the Kansai dialect, they associate their speech with the Osaka variety, which carries a nontraditional feminine image. These responses also reveal another aspect of the Kyoto variety: while the Kyoto
variety is ideologically established as a consistent image as (Figure 1.5) shows, it is not enregistered (Agha 2007) as actual audible speech. This is why one participant did not know how Kyoto speech sounded, and many identified their speech with the more highly enregistered speech variety, the Kansai dialect.

In contrast, the Osaka variety is highly enregistered, not only ideologically but also as an actual speech variety. The consistently positive responses to the question if they speak the Osaka variety provide good evidence that the Osaka variety is an enregistered speech variety. Additionally, to the question, “Are you proud of speaking your own variety?” the only two participants among all the 46 participants who responded that they felt proud of their regional language variety were both from Osaka. In addition two of the other 14 Osaka participants who did not respond positively told me that their classmates from other regions were excited to hear their Osaka speech when they first met. One also said that when she took a taxi in Tokyo, the driver knew she came from Osaka right away because of her speech. These reactions that the Osaka participants received from non-Osaka people clearly show that the Osaka speech is widely acknowledged in the society as a whole.

To the same question, “Are you proud of speaking your own variety?” none of the Kyoto participants’ responses was positive, and the most common reason was simply that they had never thought of their own variety very deeply. One Kyoto participant who regarded her variety as the Kansai dialect noted that she intentionally used the Kansai dialectal features to show her localism when she visited Tokyo. Her comments also indicative of her assumption that the Kansai dialect would be recognized in Tokyo since it is highly enregistered. Responses from the Tokyo participants presented very similar discourse to one another. All the Tokyo participants responded negatively by saying that they did not feel proud of speaking “hyōjungo.” They commonly expressed their envious feelings toward people from other regions having “unique” dialectal features. In the responses, the Osaka variety was repeatedly mentioned as an example
of a unique regional variety with specific phrases that, they think, represent the Osaka variety (e.g. *nande-yanen* ‘Why the heck?’ which is a common phrase used by Osaka comedians), but the Kyoto variety was rarely brought up by the Tokyo participants. Again, it is conspicuous that the Osaka variety is highly enregistered and recognized with a clear image of how it actually sounds, whereas the Kyoto variety is not.

1.6.4. Metalinguistic Commentaries and Japanese Women’s Language

Questions regarding self-evaluations of the participants’ own speech included the topic of JWL. After the topics of regional language varieties, the interview questionnaire (see Appendix II in Chapter 2) listed questions about JWL: “Do you think you use JWL in daily life? (If not) Why don’t you use it? (If yes) Do you use it all the time?”; “If anything, in which situations do you use it?”; and “What kind of images do you have toward JWL or women who use JWL?”

Before asking these questions on the questionnaire, I first confirmed that all the participants were aware of the terms and the language variety, with the question “Have you heard of something called *joseego* (‘women’s language’) or *onna-kotoba* (‘women’s language’) in Japanese?” *Joseego* (‘women’s language’) and *onna-kotoba* (‘women’s language’) are both commonly used in public discourses as well as JWL literature. All of the participants expressed their awareness of the terms, variety, and ideology by mentioning a few specific JWL linguistic features. Among the features they mentioned, the most commonly referred to feature was phrase-final particles (e.g. “Is it like ‘-dawa’ and ‘-noyo’?”).

The image of *otherness* becomes apparent from the responses to the questions on JWL. Regarding whether or not they use JWL on a daily basis, none of the 46 participants gave a positive response. Their reasons for not using JWL included their “not being very girly” and that they “behave naturally,” while many could not provide me with any particular reasoning since
they “have not thought of it before.” One participant also noted that “men and women are equal in the younger generations” as a reason for not using JWL. A few participants from Kyoto and Osaka additionally explained the absence of the JWL features in their speech based on the fact that the women’s language is not part of their local varieties. To the question about any special occasions they would use JWL, several participants were equating JWL with the polite speech register and responded that if anything, they might use JWL when talking to superiors or strangers.

These responses about their use/nonuse of JWL also link to the images of JWL described by the participants. The participants were further asked to elicit images they have toward JWL by answering the questions “What kind of images do you have toward the women’s language?”; “What kind of women do you think would use the women’s language?”; and “What kind of images do you have toward women who use the women’s language on a daily basis?” The word cloud in Figure 1.8 includes all the descriptive adjectives and nouns used in the responses. Most of the participants repeated the associated images they described in the responses to the previous few questions (“Why don’t you use JWL?” “If anything, in which situations do you use it?”) and further elaborated the images.
The image of JWL is consistent across the regions. *Burikko* is the word that most frequently appeared in the commentaries (12 responses) regarding the image of JWL. *Burikko* is a term invented and developed in the 1980s to refer to women who display “bogus innocence” in the performance of ultra-femininity (Miller 2004b) simply to attract men. Though *burikko* was considered a natural product of femininity in the 1980s (Miller 2004b), the term is used exclusively in a deprecatory or ridiculing manner today. A few participants commented that they would use JWL “jokingly (*jookude*)” with their friends. Other participants described JWL as “fake (*tsukutteru*),” “girly (*onnanoko-onnanoko shiteru*),” “cute (*kawaii*),” and “annoying (*iratto suru*),” which are all also commonly associated with *buriko*. For example, bringing up the name of a popular female idol as an example, one participant remarked, “When I heard [the idol] saying ‘I can’t help using [the honorific-prefix] –o before words,’ I was so annoyed thinking she’s totally faking it!” Another participant said, “At a group blind date, as soon as guys arrive, I
see some girls quickly change the tone (pitch) of their voice, and that’s when I notice that they are doing burikko!”

Other participants, while showing negative judgments toward burikko, agreed that JWL is a good “dating language” because it gives a good impression to men. JWL as a “dating language” stems from the “idealistic” figure of young women called “ojo-sama,” who is also a stereotypical user of JWL. Ojo-sama refers to a daughter from a wealthy upper-class family, who was brought up with tender care in isolation from the cruelties of the real world (M. Inoue 2006: 202). Figure 1.8 also presents other associated terms to this concept of ojo-sama such as “good upbringing,” “classy,” “polite,” and “beautification.”

1.7. Current Study

1.7.1. General Research Questions of this Dissertation

These contrastive local stereotypes and the participants’ self-images of their own speech are the starting point of this dissertation. Considering these background facts, three general research questions of this dissertation are raised:

1) How do women in these regions construct femininity using varieties with different femininity rankings (as shown in Figure 1.2)? What kinds of femininities are constructed?
2) How do women’s linguistic practices in these regions correspond or fail to correspond with the regional stereotypes of the varieties?
3) While the Kyoto and Osaka varieties share many linguistic features and are categorized under the same dialect region called the Kinki region, do the three phonetic features (breathy voice, acoustic characteristics of fricatives, and accented or deaccented intonation of negative polar
questions; which have been suggested to carry some gender associations) present the same patterns in Kyoto and Osaka?

1.7.2. Linguistic Features

To attempt to answer the research questions, I look at three phonetic features: breathy voice, acoustic characteristics of fricatives, and accented or deaccented intonation of negative polar questions. These phonetic features were chosen for the investigation of this study based mainly on three reasons. 1) Previous research on Japanese and other languages has noted possible associations between these features and gender. 2) These features in Japanese remain understudied in sociolinguistics and dialectology (see Chapters 3-5 for overviews of previous research on each feature). 3) These features are not part of the traditional JWL ideology and therefore, are free from gender meanings predetermined by the ideology. As we have seen above, the JWL ideology considers a number of linguistic features, such as phrase-final particles, self-referential terms, or honorific expressions, to mark gender, but this study will illuminate that Japanese women utilize non-JWL features to construct their gender identities. Previous studies (e.g. Okamoto 1996, Miyazaki 2002, 2004, Matsumoto 2004, Miller 2004a) have shown that these JWL features are not widely used in real women’s linguistic practices, especially among young women, which further suggests that there are other linguistic resources for gender construction. Below I provide brief descriptions of the three phonetic features and their parameters.

• **Breathy voice** refers to a voice quality type with audible frication caused by persistent air leakage from the glottis. Previous research, including a few studies on Japanese, has consistently reported a correlation of the female gender and higher degrees of breathy voice.
Degrees of breathiness will be discussed based on spectral metrics such as H1-H2, H1-A1, H1-A2, and H1-A3.

• **Voiceless sibilant fricatives** /s/ (e.g. *sumi* ‘charcoal’, *kasa* ‘umbrella’) and /ɕ/ (e.g. *shashin* ‘picture’, *shumi* ‘hobby’) will be investigated in terms of their acoustic characteristics, determined by measuring the spectral peak and the four spectral moments associated with the energy of the frication (center of gravity (COG), standard deviation (SD), skewness, and kurtosis of the frication energy). While many studies of English fricatives have found that women’s speech exhibits higher frequency energy, such an association has not been reported in Japanese.

• **Intonation of negative polar questions** (e.g. *amaku nai?* ‘isn’t [this] sweet?’ *onnaja nai?* ‘isn’t [this] a woman?’) is concerned with whether accentual phrases (APs) in a negative polar question (or a ‘yes/no’ question) retain the lexical accent of the main content word. The form where the accent is retained (the accented form) follows the conventional pattern, while the form where the accent is deleted (the deaccented form) is a newly emerged pattern in Tokyo since the 1990s (Tanaka 2010, Inoue 1998). A previous survey study (Tanaka 2010) showed that this feature carries the meaning of “urban cuteness” among young people in the Tokyo area, but the use of this feature in other regions was not previously investigated.

For the analysis of these linguistic features in Japanese, I will apply a sociolinguistic variationist methodology. As is seen in its history presented above, Japanese Dialectology has developed its own methodological and analytical approaches that are independent from the Western sociolinguistic variationist approaches. One of the gaps between Japanese Dialectology and
Western sociolinguistic variation is the discussion of social meanings (the stances, personal characteristics, and personas associated with use of linguistic forms; social meaning will be discussed in the next section). Though researchers of Japanese Dialectology have studied a number of linguistic variables from lexical items (e.g. Yanagita 1927, Tachibana 1936, NLRI 1966), pitch-accentual patterns (e.g. Hattori 1930, Kindaichi 1964), phonological features (e.g. Tsuzuku 1949, Kindaichi 1964, Kato 1975), and morphosyntactic elements (e.g. Tojo 1954), there has been virtually no discussion on social meanings of those variables (beyond their association with various geographic regions). In Western sociolinguistic variation studies, researchers have devoted increasing attention to the social meanings of linguistic variables since the emergence of the so-called “third wave” of variation study (Eckert 2012). The following section discusses the notion of social meaning in third wave variation study, while briefly reviewing the other two waves of research. As Eckert (2005) emphasizes, it should be noted that the numbering of these waves does not indicate their historical order and that three waves are not clearly separable; each wave represents a specific way of approaching variation and a specific methodological and analytic practice. While I align my study with third wave variation study, particularly in the approach to the social meaning, this study can also be classified under all three waves in terms of the methodological and analytic approaches.

### 1.8. Social Meaning

#### 1.8.1. Three Waves of Sociolinguistic Variation

According to Eckert (2012), the first wave of sociolinguistic variation studies, initiated by Labov’s (1966) New York City study, treats linguistic variants as markers of primary social categories such as socioeconomic class, gender, race, and age. This research involves large-scale surveys of geographically defined urban populations and examines intra-speaker variation by
drawing distinction between casual and careful speech. This survey method enables researchers to collect large volumes of data and create replicable studies. The social meaning of individual variants is interpreted based on their correlation with predetermined *global* categories (e.g. socioeconomic class, gender, region).

The second wave of sociolinguistic variation (e.g. Labov 1963, Milroy 1980) focuses on ethnographic work in smaller communities for relatively long periods of time in order to discover locally salient social categories and investigate how ways of speaking are imbued with local meaning. The second wave establishes a connection between local categories to the big picture of the first wave, and treats variables as indexing locally defined categories. Through ethnographic observations, second wave researchers (e.g. Eckert 1989) turn more attention to the non-linguistic aspects of identity practices in addition to observing linguistic practices in the construction of speakers’ identities.

While the second wave is restricted to fairly limited geographic entities, third wave variation studies link individual speakers’ day-to-day experience to the larger community of practice. Rather than treating linguistic variants merely as markers of primary social categories or locally defined categories, third wave variation studies seek to uncover interactional as well as group-associational meanings attached to each variant and how those variants are placed as integral parts of distinctive ways of speaking. Thus, the social meaning of linguistic variation plays a critical role in the construction of personal styles or persona. As Eckert (2005: 24) notes, “the meaning of variation lies in its role in the construction of styles, and studying the role of variation in stylistic practice involves not simply placing variables in styles, but in understanding this placement as an integral part of the construction of social meaning.” Style will be discussed at length in Chapter 6 of this dissertation.
1.8.2. Indexical Meaning

Throughout this dissertation, I make use of the theoretical concept of indexicality to link linguistic practices to social information. The term “index” used in sociolinguistics carries a broad meaning and encompasses C.S. Peirce’s three classifications of signs. According to Peirce (1931-58), signs (or “representamens”) are meaningful as either index, icon, or symbols. As articulated by Coupland (2007: 22), in Peirce’s narrow definition, an index refers to “a relationship between a sign and a referent (the object that it is linked to) which is based on a physical or in some other way objective or ‘real’ association,” and an icon is distinguishable from an index in that the iconic link is based on “natural resemblance” between a sign and a referent. A symbol is also different from an index since it involves a social process which establishes arbitrary and conventionalized link between signs and meanings. In socioinguistics, many well-studied phonological variables such as postvocalic r-lessness carry symbolic meanings since the association between the feature and vernacularity or “nonstandardness” (or many other meanings like African American ethnicity and Southern locality; cf. Schilling-Estes 2004) is arbitrary rather than “real” or “natural” in any sense.

With a broader definition of “index,” Silverstein (1976: 30) describes two kinds of indexes, namely “referential indexes” (or shifters) and “non-referential/pure indexes.” Referential indexes, including deictics such as demonstratives, pronouns, and tense markers, contribute to the referential content of discourse but also depend on context. For instance, the English pronoun I has referential content since it refers to the first person singular but the meaning depends on who the speaker is. On the other hand, non-referential indexes point to aspects of the context that are apart from the referential content of speech such as gender, region, and stances. This type of indexicality of speech connotes social meanings through which people construct and understand identities.
Anthropologists Silverstein (1985, 2003) and Ochs (1992) demonstrate that the interpretation of social meaning can be observed at different levels. As Hall-Lew (2011) notes, this “indexical order” is one of the most confused (and confusing) areas in the current sociolinguistic literature. Silverstein proposed an $n^{th}$-order or an $n+1^{st}$-order meaning, but scholars generally concentrate on first and second order indexicals since the two levels present very different semiotic processes which link a sign to a meaning. This interpretive schematization begins with one correlation at the $n^{th}$-indexical, and the correlation becomes reinterpreted through use in interaction creating newer meanings at the $n+1^{st}$- or higher order indexicals (Hall-Lew 2011). For example, the pronominal alternations in Japanese between omae ‘you’ and anata ‘you’ gain different first-order indexical meanings. Omae is used among friends or people with higher status addressing the lower, while anata is used among strangers or by people with lower status addressing the higher. These correlations characterize omae as casual/deprecatory and anata as formal/respectful at the level of first-order indexicality.

Second-order indexicality involves the “cultural construal of the first-order indexical association with an intentional content or meaning” (Woolard 2008: 437). At this second level, “actors rationalize, explain, and thus inevitably naturalize and ideologize the sociolinguistic associations (indexical relations) that they have registered at the first order” (Woolard 2008: 437-38). At this level, the use of omae may start being interpreted to indicate the speaker’s affect such as aggression, which ideologically derives from the first-order meaning (i.e. deprecatory).

A special type of indexicality, the notion of iconization, discussed by Irvine and Gal (2000), holds a similar function to Silverstein’s second-order idexicality. Iconization is “a semiotic process that transforms the sign relationship between linguistic features and the social images to which they are linked” (Irvine 2001: 33). Through iconization, people treat social images indexed by linguistic features as natural or essential qualities of the features or the speakers. These links are usually constructed through “metapragmatic practices” (Silverstein
1993). In the example above, people who use *omae* may be perceived as aggressive and tough through iconization. Here, its original referent (the actual semiotic process; i.e. indexicality of hierarchy) gets lost, and the feature itself starts offering social meanings. In other words, iconization is the process whereby the link between a linguistic form and its meaning is made to look non-arbitrary, and in Coupland’s (2007) sense, symbols are transformed to appear as icons and indexes through this process (“*Omae* naturally means tough” or “*Omae* is a tough word.”).

What makes the notion of indexical order confusing is that the interpretations of first- and second-order (or $n^{th}$ to $n+1^{st}$ order indexicality) indexicalities vary depending on the researchers. Some sociolinguists (e.g. Woolard 2008, Foulkes, Scobbie, and Watt 2010) interpret the distinction between first-order and second-order indexicalities depending on whether the link is “pre-ideological” or “post-ideological,” but others (e.g. Moore and Podesva 2009) question the distinction. Some researchers (e.g. Johnstone and Kiesling 2008, Johnstone 2009) conceptualize the order as a linear semiotic development, while others (e.g. Eckert 2008b) argue that they are mutually and simultaneously constitutive. In Foulkes, Scobbie, and Watt’s (2010: 738) interpretation, “[f]irst-order indexicality refers to the (objective) association of particular patterns of linguistic behavior with globally or locally meaningful social groups, while second-order indexicality pertains to speakers’ subjective metalinguistic knowledge of the social and communicative roles played by variable linguistic forms.” Eckert (2008b: 28) argues that “[f]irst order indexicality renders the linguistic feature in question available for association with stereotypes associated with the category. The minute such an association materializes in practice – as soon as speakers begin to use a feature to signal something associated with the category – the feature becomes a second order index.”

According to Johnstone and Kiesling (Johnstone and Kiesling 2008, Johnstone 2009), the indexicality of Pittsburgh’s local variants (e.g. ‘downtown’ pronounced as *dahntahn*; *gumban* meaning ‘rubber band’) shifts from $n^{th}$ to $n+1^{st}$ order indexicality over time due to
exposure to forms in other regions through social and geographical mobility. In their interpretation of Silverstein (2003), first-order indexicality shows only a correlation between local variants and the speakers’ demographic facts, such as whether the speakers are from Pittsburgh and are working-class male. At this point, there are no metapragmatic practices regarding the first-order correlation, and there are no ideologies around the regional forms. At the level of second-order indexicality, “speakers start to notice and attribute meaning to regional variants and shift styles in their own speech.” The meanings of the regional variants are shaped by ideologies such that the forms index working-class identity, incorrectness, and lack of education. At the third-order, indexical meanings of the forms shift, and people who are aware of second-order stylistic variation in Pittsburghers’ speech link the regional variants to Pittsburgh local identity. Some begin to use the variants strategically to express their local pride, “drawing on the idea that places and dialects are essentially linked” (164).

Although Johnstone and Kiesling treat indexical order as linear and temporal, Eckert (2008a) points out that this interpretation is not Silverstein’s intention. Silverstein (2003: 194) states that the $n+1^{\text{st}}$ order is “always already immanent” in the $n^{\text{th}}$ order; similarly, Eckert argues that the indexical links co-exist by taking place within “a fluid ever-changing ideological field” (464). Hall-Lew (2011) acknowledges that the term ‘order’ may be part of the cause of the confusion and clarifies that ‘order’ is “in the sense of ‘social order’, referring to relations among elements of a system.” Furthermore, as noted above, even though Silverstein describes $n^{\text{th}}$ order meanings as “pre-ideological” and $n+1^{\text{st}}$ order meanings as “post-ideological,” the distinction is questioned by Moore and Podesva (2009) since “there is no possible absolutely pre-ideological – that is, zero-order, social semiotic” (Silverstein 1998: 128-129; cited in Moore and Podesva 2009: 450). Therefore, Moore and Podesva suggest that researchers consider the relationship “with a different level of social meaning in terms of an $n+1^{\text{st}}$ order indexical ideologization”
(Moore and Podesva 2009: 450) rather than attempting to distinguish pre-ideological and post-ideological phases.

While Silverstein’s model helps us to conceptualize different levels of relations between signs and meanings, Ochs’ (1992) two-tier semiotic process of indexicality is particularly useful for my investigation since it articulates how indexical processes construct identities at the interactional level. Rejecting one-on-one links between language and gender, Ochs’ model accounts for the fact that some types of social meaning are more directly indexed than others. A linguistic form, including lexical, syntactic, phonological, or any other form, does not have a direct relationship to a social category; instead, a social category is mediated by “stances, social acts, social activities, and other social constructs” (Ochs 1992: 337), which in turn are associated with social categories. At this level of indirect indexicality, ideology comes most centrally into play, since stances acquire more enduring semiotic associations.

For example, in Ochs’ model, what is considered a Japanese feminine sentence final particle such as -wa does not directly index femininity. Rather, its direct index of a gentle stance indirectly indexes femininity. Moreover, as Bucholtz (2009: 148) points out, linguistic forms that may come to be ideologically linked to broader social meanings via indirect indexicality are generally associated not with broad social categories like women but rather with more specific sorts of social types and personas through the process of creating metapragmatic stereotypes. Metapragmatic stereotypes are defined as “culture-internal models of utterance indexicality associated with speech variants” (Agha 2007: 148). Hence, more specifically, the particle -wa indirectly indexes the ‘good wife, wise mother’ figure. Over time, through repetitive uses of the feature by specific types of people in specific contexts, the mapping between linguistic form and social meaning (or specific persona) comes to be ideologically perceived as direct, and the distinction between the direct and indirect indices may be lost.
1.8.3. Indexical Field

Building upon the notion of indexicality, Eckert (2008) proposed what is called indexical field. Unlike Silverstein and Och’s theories, indexical field is not concerned with different levels of indexicality. Instead, multiple potential meanings of a linguistic variant and ideologically related meanings are constituents of the field where any one(s) of the meanings can be activated in the situated use of the variant. This model offers representations of the notion that has been established by the collection of previous studies that social meanings of linguistic variables are neither precise nor fixed, but rather, they are multifaceted and fluid. For example, one of the most studied phonological variables, “released /t/” has been demonstrated as a resource to construct a wide range of personae, including the nerd identity of female adolescents (Bucholtz 2001), the masculinity of Orthodox Jews (Benor 2004), a “good woman” figure in Martha Stewart’s speech (Sclafani 2009), and the competence of a gay doctor (Podesva 2007). Further, Podesva et al. (forthcoming) demonstrated that released /t/ can be used by different U.S. politicians to construct various personae, but the effectiveness varies depending on the speakers. These studies clearly show diverse potential meanings of released /t/, such as being articulate, educated, elegant, prissy, and exasperated, and one or more of these meanings become more or less prominent according to the specific speech contexts and speakers.

1.8.4. Definition of Social Meaning

Though social meaning is a crucial concept in sociolinguistic variation studies, it is not necessarily defined clearly in many studies. Campbell-Kibler (2009: 136) explains that social meaning is “social content tied in the minds of a given speaker/hearer to a particular piece of linguistic behavior.” More precisely, Podesva (2011: 234) defines social meaning as “the stances,
personal characteristics, and personas indexed through the deployment of linguistic forms in interaction.” In this definition, “stance” is understood to involve an act of evaluation where “the stancetaker (1) evaluates an object, (2) positions a subject (usually the self), and (3) aligns with other subjects” (Du Bois 2007:163; cited in Moore and Podesva 2009: 448). For example, “burnouts” in Eckert’s (2000) study employ negative concord which indexes a rebellious stance built in opposition to the school (the evaluated object) and disalign themselves from their “jock” counterparts. Here, burnouts’ use of the particular feature has less to do with their identity construction per se than it has to do with the rebellious stances they habitually take.

In my study, social meaning refers to the stances, personal characteristics, and personas indexed by ideological associations in speakers’ and hearers’ minds with linguistic forms as well as non-linguistic aspects of identity practices (such as appearances and actions). Though third wave variation studies tend to focus on speakers’ creativity and agentivity (Schilling 2013a), linguistic variants do not necessarily convey certain social meanings as speakers intend (or do not intend) based simply on the use of the linguistic forms. As Campbell-Kibler’s definition of social meaning indicates, hearers are part of the semiotic process since the meanings of linguistic forms need to be interpreted and understood by the hearers who have various expectations or assumptions based on ideologies, while considering co-occurring linguistic features and non-linguistic information about the speakers. For example, Sclafani (2009) demonstrates how the meaning of the released /t/ variant used by Martha Stewart has changed after her 2003 insider trading conviction because of the drastic change in her public images. While released /t/ was associated with Stewart’s “Good Woman” figure before 2003, the very same variant is used in parodic performance to spotlight her as a “Bad Woman” after the criminal incident. In a study (Kajino 2010) of phrase final lengthening used by the female Japanese pop singer Hamasaki Ayumi, I argued that a low degree of final lengthening indicates immaturity when co-occurring with child-like features (non-polite ending forms and high mean
Fo) but means maturity when it occurs with mature features (polite ending forms and mid mean Fo).

While indexical field is very telling in that various meanings of a linguistic variant constitute a field where any of the meanings can be activated in the situated use of the variable, the fact that a given variant can have competing/conflicting meanings suggests that the meanings are not necessarily ideologically linked to one another (e.g. Sclafani 2009; Kajino 2010). The same concern holds in the current study, and so the notion of indexical field is not well suited to my present purposes.

1.9. Summary

In this chapter, I have described the dialectal images of the three regions and showed how the particular images construct distinctive gendered stereotypes, such as the boring, standard, but cute Tokyo figure, the classy and calm geisha Kyoto figure, and the harsh and funny Osaka Auntie figure. These stereotypes are by-products of the interaction between region and gender, so this chapter has provided overviews of Japanese Women’s Language and Japanese Dialectology as the two relevant fields of this dissertation. I have also outlined the central theoretical and methodological approaches of the study. Particularly, the concept of social meaning has been highlighted with a brief review of the third wave of variation.

1.10. Organization of the Dissertation

The following chapter, Chapter 2, will describe the data of this dissertation and data collection methods. I will outline the basic background of the three women’s colleges, the research sites, in
terms of their histories, demography of the students, characteristics, and communities of practice.

Chapters 3-5 will each analyze one linguistic feature using the data described in Chapter 2. Chapter 3 focuses on the use of breathy voice by the participants across the three regions. This chapter demonstrates that a linguistic feature that is usually associated with gender like breathy voice can also present regional variation. The analysis shows that Kyoto is significantly different from the other two in the use of breathy voice. While the Tokyo and Osaka participants use relatively consistent degrees of breathiness regardless of different levels of intensity and pitch, degrees of breathiness of the Kyoto participants’ speech strongly correlate with low intensity and high pitch, which helps Kyoto women construct the traditionally feminine figure.

Chapter 4 will turn our attention to the realization of fricatives /s/ and /ɕ/. While Chapter 3 found that a feature that carries gender meanings presents regional differences, this chapter reveals that a feature that distinguishes regions can simultaneously connote social meanings that can be used in gender construction. To understand meanings of variants, this chapter borrows some psychoacoustic notions of sound symbolism such as sharpness and clarity. The parameter of sharpness well describes the Osaka participants’ use of fricatives with higher frequency energy, shown with higher COG and lower skewness. On the other hand, with a lower SD and higher kurtosis of the frication energy, the fricatives produced by the Tokyo speakers exhibit clear and precise articulations. These sound qualities of fricatives further correspond with the stereotypical images of the regional varieties.

In Chapter 5, I will investigate the use of the deaccented form of negative polar questions (NPQs) and explore discursive roles and the social meaning of the feature in construction of femininities and locality. This chapter will demonstrate how a feature that carries the meaning of the Tokyo regionality (i.e. deaccented NPQs) can be used by non-Tokyo speakers for their
gender performance. Quantitative analysis reveals cross-regional patterns, and qualitative analysis shows differing meanings of the feature across the regions.

Chapter 6 will explore the meanings of breathy voice and height of frequency energy of /s/ by locating the variants in the participants’ stylistic practices. This chapter will highlight intra-regional diversity through interspeaker variation in the participants’ styles. While some participants align themselves with local stereotypes by using local linguistic features, others disalign themselves from the local stereotypes by avoiding those local linguistic features.

Finally, Chapter 7 will summarize the present study’s findings on Japanese women’s linguistic practices across Tokyo, Kyoto, and Osaka and discuss how this dissertation fits in the two relevant fields of Japanese Women’s Language and Japanese Dialectology. Limitations of this study will also be noted for future direction of sociophonetic studies of Japanese language, gender, and region.
CHAPTER 2

Fieldwork

2.1. Introduction

As is encouraged by Schilling (2013b), I utilize multiple data-collection techniques for this dissertation in order to take advantage of strengths and ameliorate shortcomings of different methods. The main data on which the analysis of this dissertation is based consists of over 40 hours of 25 recordings of conversations with 46 female college students between the ages of 18 and 21 years old. The 46 participants include 15 students from Tokyo, 15 students from Kyoto, and 16 students from Osaka. The spontaneous conversations were recorded with pairs of two close friends without the researcher’s presence. The participants were personally recruited at Tokyo Rose University, Kyoto Cherry University, and Osaka Lily University (pseudonyms). In addition to these recorded conversations, I conducted interviews with each pair of participants focusing on their demographic backgrounds and beliefs and ideologies about language and gender. A mini-ethnography was also conducted in order to become familiarized with the participants’ communities of practice.

In this chapter, I will first discuss the rationale of studying the specific population of Japanese women. Then, I will introduce the research sites as well as the participants, referring to my notes from the fieldwork. I will also discuss the methods of the mini-ethnography, conversation recording, and interviews.

2.2. Research Sites

2.2.1. Rationale
The research sites are three private women’s universities in the Tokyo, Kyoto, and Osaka prefectures: Tokyo Rose University, Kyoto Cherry University, and Osaka Lily University. The particular age group of university students was selected since this study is concerned with Japanese women’s gender construction as well as with their negotiations through regional stereotypes and social pressure regarding gendered ideology. This age group, being in transition between adolescence and adulthood, has characteristics of both adolescence and adulthood. Adolescence is an age group that has received much scholarly attention in sociolinguistics since the 1980s because of linguistic innovativeness and vernacular speech. Sociolinguists generally agree that adolescence is the life stage in which language change is most clearly visible (Kerswill 1996) and greatest inter-speaker variation can be observed (Eckert 2004). In addition, adolescents are often found to prefer local vernacular varieties and variants more so than people in other age groups (e.g. Labov 1972, Eckert 1995, Kerswill 1996).

While these characteristics of adolescence are advantageous in studying regional linguistic usages and innovative gender construction, negotiation through social pressure of the dominant gendered ideology is not a great part of adolescents’ daily experience. What defines adolescents linguistically is their devotion to producing differences, whereas in adulthood, people conform more to the existing social order (Eckert 2004). Obvious counter-culture and social deviations can be a powerful driving force for adolescents’ linguistic and non-linguistic gender practices, but post-adolescents, through exposure to the broader society outside of school, learn that those differences have to be produced in a somewhat subtler manner. In Miyazaki’s (2002, 2004) studies of junior high school students, it is reported that some female students use the masculine self-referential forms such as boku and ore to present rebelliousness. These girls are most likely to terminate this obvious deviation of pronominal use before they reach adulthood or post-adolescence. Then, without the obvious marking like the unconventional use of self-referential forms, how do they produce the distinction? Being in-
between adolescence and adulthood, post-adolescent young adult females showcase less obvious but still very complex gender construction.

To collect the speech data of post-adolescent young adults, universities are great sites. Women’s universities are more suited for this study than mixed-sex universities since they tend to be more tied to the surrounding local communities than mixed-sex universities. While many mixed-sex universities focus on their cosmopolitan image in their marketing, women’s universities tend to attract students with their local images. These differences in marketing strategies are a result of mixed-sex universities being much larger in size than women’s universities. In fact, the smaller scale of women’s universities was another reason why I chose women’s universities. The small-scaled campuses made fieldwork manageable.

The strong ties to the local communities that women’s universities tend to maintain are also rooted in their historical backgrounds. There are many women’s universities in Japan while there are no men’s universities to my knowledge. This is because many women’s universities were established as educational institutions for local women during the time period when women were not allowed into regular universities. Even long after those regular universities became mixed-sex universities by accepting female students, women’s universities today still remain and continue serving as local institutions for higher education. Most Japanese women’s universities, other than a few national and prefectural universities, are privately funded, and the three research sites of my study are not exceptions.

Students of women’s universities in Japan compose a well-established imagined community (Anderson 1991), called joshidaisee ‘women’s college student,’ where it is believed that members of the community share gendered ideologies and practices. The imagined community is often featured in mass media such as fashion magazines, which represent young adult women in a feminine, conservative, and luxurious fashion who strive to find future
husbands (see discussion on the *JJ* magazine in Kajino and Podesva 2007). It is merely an imagined community and not a reality, but the ideologically loaded figure of a typical *joshidaisee* is so highly enregistered that it is inevitable for those students to position themselves relative to the figure – whether it is convergence or divergence – in their gender identity consideration. This makes their gender construction somewhat comparable across different geographical regions.14

### 2.2.2. Tokyo Rose, Kyoto Cherry, and Osaka Lily

From many Japanese women’s universities in the three regions, Tokyo Rose University, Kyoto Cherry University, and Osaka Lily University were chosen as research sites because of their similarities in size, history, offered curricula, and the overall academic levels. Tokyo Rose University was founded in 1987 based on Christian principles, following the establishment of an associated two-year college, and today, the university has about 1,650 undergraduate students. Kyoto Cherry University, founded as a Buddhist university in 1964, has about 1,500 undergraduate students. Osaka Lily University has approximately 1,500 undergraduate students today and was founded based on a Christian mission in 1964, following the establishment of its two-year college. The universities in Tokyo and Osaka are Christianity-based, whereas the university in Kyoto is Buddhism-based. Buddhism-based educational institutions have strong roots in Kyoto while Christian institutions are more common in Tokyo and Osaka. Though a professor at Kyoto Cherry University told me that there are a few students who are daughters of Buddhist monks at the university, a majority of students at these three universities are not

14 In fact, the participants often talked about themselves as aligning with or disaligning from the stereotypical *joshidaisee* figure in their recorded conversations. For example, a few participants noted that the conversations they were carrying out were atypical of *joshidaisee*-like conversations. These comments show that the *joshidaisee* figure provides, in a way, a reference for their gendered self-images, including purposefully orienting away from the stereotypical image.
affiliated with any religious group as a majority of the Japanese population is not religious.

None of the participants in my study claimed any religious beliefs.

<table>
<thead>
<tr>
<th></th>
<th>Tokyo Rose</th>
<th>Kyoto Cherry</th>
<th>Osaka Lily</th>
</tr>
</thead>
<tbody>
<tr>
<td>N of students</td>
<td>1,650</td>
<td>1,500</td>
<td>1,500</td>
</tr>
<tr>
<td>Year founded</td>
<td>1987</td>
<td>1964</td>
<td>1964</td>
</tr>
<tr>
<td>Religion</td>
<td>Christianity</td>
<td>Buddhism</td>
<td>Christianity</td>
</tr>
<tr>
<td>Majors</td>
<td>Japanese, English,</td>
<td>Japanese, English,</td>
<td>Japanese, English,</td>
</tr>
<tr>
<td></td>
<td>Psychology, Social</td>
<td>Social Science /</td>
<td>Social Science,</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>Psychology, Nutrition</td>
<td>Psychology, Nursing</td>
</tr>
</tbody>
</table>

Table 2.1. Summary of Research Sites

The three universities offer a limited number of majors, which are comparable to one another. The universities all have departments of Japanese language and culture, English communication, social sciences, and psychology. Additionally, Tokyo Rose University offers a unique program in horticulture under the social science department, Kyoto Cherry University offers programs in nutrition, and Osaka Lily University offers a program in nursing. While the two-year colleges at Kyoto Cherry University and Osaka Lily University are still open, Tokyo Rose University closed their two-year college in the early 2000s. Students in these two-year colleges are not subjects of my study.

There are also affiliated junior high schools and high schools for the three universities though they have separate campuses in different locations. The system of affiliated institutions from secondary or even primary education to higher education is common among private schools in Japan, especially among women’s institutions. Usually, once students pass the entrance examinations and matriculate in the school at the secondary or primary levels, as long
as they stay in a particular school, they can advance to the upper levels with “preferential treatment” without taking official entrance examinations. This educational system is commonly described as “escalator-style” in a slightly cynical manner because of the privileged image. In general, the tuition for private educational institutions is two to three times more than tuition at national or public institutions, and the scholarship system is not very common for Japanese students. The “escalator-style” is convenient to students, but colleges with this system tend to be seen as academically less competitive than others that require entrance examinations to screen students who wish to advance into the upper levels. Therefore, in reality, many students at the affiliated high schools of my research sites leave the institutions and enter larger universities. Every year Tokyo Rose, Kyoto Cherry, and Osaka Lily experience general turnover of the students at the college level. Tokyo Rose University has only a few or no students from Tokyo Rose High School each year, while Kyoto Cherry University and Osaka Lily University welcome about 50 students from the affiliated high schools.

These universities are each located within 40 minutes by the metro system from the centers of the cities – Shinjuku in Tokyo, Kawaramachi in Kyoto, and Umeda in Osaka – which are often mentioned by students as popular destinations for shopping, dining, and having fun. Among the three universities, Tokyo Rose University is farthest from the city center, which makes the university more locally tied compared to other universities in Tokyo which attract students from other areas. The list of high schools from which the first-year students in 2011 matriculate, published by Tokyo Rose University, shows that the majority of students are from Tokyo and the adjacent Kanagawa Prefectures.

The commentaries on the images of the three universities acquired from the students and people in the communities during my fieldwork also have similarities. Students at the three universities described their schools using common keywords like “peaceful,” “easy,” “slow,” “small,” and “kind teachers.” University employees as well as small business owners in the
neighborhoods commonly mentioned how the student bodies have changed over time. The three universities held the reputation as educational institutions for ojo-sama until a few decades ago. The term ojo-sama refers to “a daughter from a wealthy upper-class family, who was brought up with tender care in isolation from the cruelties of the real world” (M. Inoue 2006: 202). However, they think that today’s students do not seem to fit the image of ojo-sama. For example, an administrative employee at Tokyo Rose University complained about some loud-speaking and ill-mannered students. The owner and customers of a small coffee shop in the neighborhood of Osaka Lily University mentioned that today’s students seem to dress like gyaru not ojo-sama. The term gyaru, derived from the English term ‘gal’, refers to a style culture that is identified with heavily dyed hair, decorative nails, and dramatic make-up. In fact, many students of the three universities enjoy hair coloring as well as nail art. It was very common among the students to wear artificial or extension eyelashes, and their makeup endeavors include wearing colored contact lenses just to change the size and colors of their irises. A long-term professor at Osaka Lily University also told me that until 20 years ago, their students’ fashion styles were featured in almost every issue of female fashion magazines specializing in conservative feminine styles, but today, they are no longer in those magazines since popular styles have shifted and become less conservative and less feminine. Finally, a professor at Kyoto Cherry University thinks that students at his university dress rather modestly and simply, but several Kyoto Cherry students I met noted that there are “many gyaru and heavy smokers” among students, but that those students tend not to come to the campus very often.

2.3. Initial Contact with the Research Sites

Fieldwork was conducted from May to July in 2011 in Kyoto and Osaka and in November and December of 2011 in Tokyo. Since the universities in Kyoto and Osaka are connected by public
transportation within an hour and a half, I conducted the fieldwork during the same time period. While I received permission to conduct research at Rose University through my personal contact, I needed to find other universities in Kyoto and Osaka on my own. For the initial selection of research sites in Kyoto and Osaka, I first determined target universities based on their location, academic level, and size as I explained above. I looked up the target universities’ websites and identified one professor at each university with somewhat relevant specializations in sociolinguistics. I contacted them via email four months in advance with a brief description of my study and asked them if I could spend a few months at their universities to conduct data collection. The professor at Kyoto Cherry University unilaterally consented to my request.

To obtain permission to conduct research at Osaka Lily University, an approval from their ethical review board was required. The review board was particularly careful with the interview questions I was planning to ask the participants. To protect the participants’ anonymity, questions that identify areas within Osaka where they reside or have resided were removed. Questions determining the participants’ socioeconomic classes also needed to be removed since it is a sensitive subject in Japanese society. Furthermore, regarding my research design, which involved explaining to the participants that this research concerns their views about various issues such as food, fashion, relationships, careers, and language, rather than simply their speech, the review board initially advised that I tell them that I will analyze their speech. However, they approved after I explained that the purpose of not highlighting the focus on the language is to obtain the most relaxed conversation possible while avoiding raising their consciousness about their speech.

2.4. Mini-Ethnography
At each site, I spent the first two to three weeks focusing on a mini-ethnography. Since I graduated from a small-sized women’s college similar to the three research sites, I had a general understanding of the culture of women’s colleges. Therefore, the mini-ethnography was conducted to familiarize myself with the specific universities. I commuted to the campuses using public transportation systems as well as school buses, which is a common way of commuting for many students. I started my ethnography by visiting the admission offices to find university booklets and other materials for new students. In addition to general information about the universities that can be found on the websites, the packages included the course schedules, detailed descriptions about university facilities, and brochures on different club activities to solicit new members. Referring to the brochures, I visited various club activities to meet students from various communities of practice (Eckert and McConnell-Ginet 1992).

As Schilling (2013b: 25) notes, communities of practice are a key notion for many variationist studies focusing on gendered practices since those practices tend to be developed in day-to-day experiences. Social networks are another social construct that is concerned with local development of linguistic practices within small communities. A primary difference between the two is their foci. While the social network theory focuses on the strength of the communities’ network ties (in terms of density, clusters, and multiplexity) and speakers’ levels of involvement into the communities (Milroy 1980), the focus of communities of practice is practices shared by people in the communities, considering membership networks are deeply embedded into the practices. The approach of communities of practice is more suitable for my study since it helps us see how gender is co-constructed with non-linguistic practices (Eckert 1996).

Communities of practice on the campuses were observed through training sessions or meetings of different club activities. I learned about current issues at the universities by attending student government meetings and experienced traditional Japanese flower
arrangement with members of a flower arrangement club. I watched the training of a lacrosse team at an off-campus field and also went to a national competition for a cheerleading team. I also attended a few classes on English communication and linguistics by the professors with whom I was in contact.

I also observed students at various places on the campuses. The key to participation-observation is to develop an insider-perspective while preserving a measure of outside detachment (Schilling 2013b: 113). Especially, I spent much time at the university cafeterias and on-campus coffee shops. Not only do students dine there, but they also pass time between classes or appointments and relax there. Chatting with friends and fixing makeup are common activities that take place in these spaces. Some students play with portable game consoles, and some others read mangas (‘comics’). Studying and taking naps were also commonly observed. A few less common activities include card games, loud singing, and arm-wrestling.

While being at the cafeterias and coffee shops, I also talked to many students. Most of the students I spoke to were willing to talk to me. Though it may sound more challenging, I realized that approaching a bigger group of friends is likely to be more successful especially among the first or second year students. It is probably because they feel “safer” if more friends are involved together, while a group of only two friends tended to ask each other to agree to talk to me. If the students are juniors or seniors, they usually have some ideas about conducting research and also are more accustomed to talking to a strange older person, so communication tended to go more smoothly. To initiate conversations, I introduced myself as an outside student who was conducting research on the campuses and was interested in students’ opinions about their universities as well as the neighborhood. After asking a few questions about their impressions of the universities, I usually shifted to slightly more personal questions like where they go to have fun and what they like to do in their free time. These questions are informative about the students and their culture and useful for building rapport with them. Once the
students felt comfortable with me, they told me about various issues, including personal and sensitive matters.

2.5. Participants in Audio Recordings

During these few weeks of mini-ethnography, I found most of the participants for the recorded conversations and interviews and scheduled recording sessions with them. The participants selected for inclusion in the conversational recording had to be students of one of the three universities who spent over two-thirds of their lives in the designated area. However, since there are likely to be more students from outside Tokyo at Tokyo Rose University, I changed the geographic parameter for Tokyo to a slightly wider area including the adjacent cities Yokohama and Kawasaki in Kanagawa Prefecture. Among 15 participants in Tokyo, 11 were from Tokyo and four were from Yokohama or Kawasaki cities. Although the initial threshold for the duration of time lived in the areas was two-thirds of their lives, it turned out to be that all the participants of this study spent their entire lives within the areas.

Before recording their conversations, I made sure to meet with each participant personally over lunch or a cup of tea for an hour or longer. These sessions were held to help myself become acquainted with each participant as well as to make participants feel comfortable with participating in my study. I did not frame the sessions as part of my research, but rather, I asked them to chat with me as a friend. Many of the participants told me about their relationship issues, and many others asked for my advice about study abroad or studying English.15 All were interested in what it is like to live in the United States, and particularly

---

15 Offering advice was one of the ways in which I gave back to the community. I met many other students than the participants in the conversation recordings to talk about study abroad, studying
interested in the topics of food, fashion, and entertainment. Some of them talked to me using polite features, and others used plain or casual forms. One of my intentions in establishing rapport with the participants was to emphasize that I would never screen or judge what they say in the recorded conversations, which is essential to obtain relaxed conversations between friends with or without the researcher's presence.

During communication with the participants, I attempted to speak varieties that are similar to the participants’. Growing up in Western Japan while being raised by a Tokyo-native mother, I consider myself bi-dialectal of the Kansai (‘Western’) dialect and the Tokyo variety to some extent. The reason for speaking similarly to the participants’ varieties was to avoid too much influence from my own regional speech on the participants’ speech and also to establish a comfortable speaking environment for them. To maintain a relaxed setting, I used the casual register, which is also appropriate considering the age differences.

A table listing pseudonyms of the 46 participants including their ages, year in university, academic department, and parents’ home prefectures is included as Appendix I at the end of this dissertation (p. 268).

2.6. Recordings of Conversations

In total, 25 conversations between pairs of students were recorded. Nine conversations by the Tokyo participants contained three speakers from outside of the Tokyo-Yokohama/Kawasaki region, and eight conversations recorded in Kyoto contained one student from outside of Kyoto. Data from these four speakers from outside of the designated regions are excluded from the English, choosing graduate programs, and finding careers using English. Professors and the conversation participants introduced many of these students to me.
analysis. All the speakers in the eight conversations recorded in Osaka were from Osaka and, therefore, are all included in the data.

The recordings were made with a digital voice recorder capable of recording two channels simultaneously (Zoom H4n) with two lapel microphones (Audio Technica AT831b) at a sampling rate of 44,000 Hz. Each participant wore one microphone so that their speech was recorded on a different channel than that of their co-participant. Lapel microphones were chosen since they are small enough not to disturb participants and allow participants to comfortably move during recordings. Each conversation lasted an hour and a half. After transcribing the conversations, all the recordings were split into two tracks. The resulting stereo sound was converted into mono sound using Audacity for acoustic analysis with PRAAT (Boersma and Weenink 2013).

Prior to recording, I obtained a consent form from each participant. I explained to the participants that the purpose of the research is to advance our understanding about young Japanese women and that I am interested in their views about any matters related to their daily lives including language. Recording took place in unused classrooms and meeting rooms on the campuses. I was absent from the room in order to minimize my influence on their speech, but for the first three conversational recordings (one in Osaka and two in Kyoto), I monitored their conversations outside of the rooms. By using a 10 meter-cable attached to the microphones on the recorder, I could listen to their conversations through headphones and monitor the sound quality. For the rest of the conversations, I used a one meter-long cable and did not monitor their conversations during the recordings. While using a cable longer than 5 meters can potentially cause sound quality loss (Schilling 2013b: 218), fortunately, the sound quality of the three conversations recorded with the 10-meter cable does not appear to be different from the others.
The method of recording pairs of friends with no researchers present was used very effectively by Stuart-Smith (1999) and Macaulay (2002a, 2002b) in their studies of speakers across different age groups, gender categories, and social classes in Glasgow. While the first set of data was recorded as sociolinguistic interviews with a single interviewer, the second set of data consisted of same-sex dyadic conversations. The participants were asked to talk about “anything they wished” with whom they can comfortably maintain conversations during the half an hour-long tape recordings (Macaulay 2002a, 2002b). According to Macaulay (2002a: 400), the resulting tapes are “free from addressee effect (Bell 1984) that might be caused by an academic interviewer.” As Schilling (2013b: 113) points out, this method was particularly effective in obtaining casual and relaxed speech data of adolescents, since adolescents may be hesitant to talk to adult researchers in the formal “interview” setting.

While Stuart-Smith (1999) and Macaulay (2002a, 2002b) did not specify conversation topics, in my study, I attempted to control their conversation topics to some extent by providing sets of magazines to look at during the recordings. Whereas addressee effect is a major concern in the field of sociolinguistic variation, topics are another factor that has been demonstrated as having a significant effect in stylistic variation (e.g. Blom and Gumperz 1972, Coupland 1981, Rickford and McNair-Knox 1994, Rickford and Price 2013, Schilling-Estes 2004). I provided the participants with four sets of current issues of magazines in different genres: 1) food and sweets; 2) women’s fashion; 3) men’s fashion; 4) jobs and careers. Within each genre, three to four different magazines were provided to accommodate the various preferences of the participants. All the women’s fashion magazines targeted young females from late teens to early 20s, varying in their styles (see Figure 2.1), while the set of the men’s fashion magazines presented men in different age groups, including college students and young businessmen, in various styles (Figure 2.2). The first set of magazines on food and sweets were used for a 10 minute-long warm-up for the participants to become used to the environment, and this session is excluded.
from the analysis. After the warm-ups, each session using the other three sets of magazines lasted 20 to 30 minutes.

![Examples of female (left) and male (right) fashion magazines used during the recordings](image)

Figure 2.1. Examples of female (left) and male (right) fashion magazines used during the recordings

Though the magazines were provided, I specifically instructed the students to talk freely in order to minimize constrictions of relaxed speech. Obtaining more relaxed conversations was a higher priority than obtaining conversations on controlled topics. I was also interested in what other kinds of topics they would bring up, which was another reason why I did not control the topics very strictly. During the actual recordings, most of the participants talked about each magazine genre as well as other topics of interest. While looking at women’s fashion magazines, conversations are often concerned with issues about their own bodies and images of “ideal” women in addition to fashion styles and shopping. Those women’s fashion magazines, published nationally, included several articles that feature some local restaurants or shopping areas in Tokyo, Kyoto, and Osaka. A few students even found their friends in an article on an aesthetic salon and the “Miss University” beauty pageant competitions. Regarding men’s fashion, many took judgmental stances toward styles and models presented in the magazines and discussed figures of “ideal” men for them. The genre of jobs and careers were more locally specific. I provided booklets of classified postings of part-time jobs in the designated areas. An educational magazine on more serious careers was also provided, but it attracted less attention from the participants.
At the end of the recording, the participants commonly commented to me that they had behaved “too normally” during the conversations since they forgot that they were being recorded. However, as Schilling (2013b) warns, it should be kept in mind that we can never eliminate observer effects completely. Even though I was out of sight while participants were being recorded, the recorder and microphones were there, and the participants were aware of the fact that I would listen to the recordings. In fact, a few participants left messages for me during the conversations such as “Thanks for bringing this many magazines for us,” and some others remarked, “Oops! She will listen to this later!” or “I feel bad that she has to listen to our silly talk.” While observer’s paradox can cause a great dilemma for linguistic researchers, Schilling (2013b: 128) concludes that rather than trying to overcome it, we should include observer effects with other contextual factors in linguistic analysis since there is no such thing as unobserved language data, even if the observers (i.e. addressees and other audience members) are not linguistic researchers.

2.7. Interviews

Upon completion of the conversation recording, I interviewed the participants. The pairs of participants were interviewed together. These interviews were recorded but not included for the data analyses of Chapters 3 to 5. Each interview lasted for about 20 to 40 minutes, and all the questions asked during the interviews are attached in Appendix II at the end of this dissertation (p. 269). The main purposes of the interviews were to collect the participants’ demographic information and comments on their regions and language varieties as well as their beliefs and ideologies about language and gender. Their views about the three regional language varieties and gender ideologies were reported in Chapter 1. I also asked questions about their communities of practice and non-linguistic practices. Questions about favorite fashion brands
and stores were added since commodity and consumption is a primary means of forming identities among young people in late capitalist societies like the United States and Japan (Bucholtz 2007), and clothing in particular plays a central role in the participants’ gender performances and stylistic practices.

I also included a question about their email addresses. The most common written communications among Japanese college students at the time of my fieldwork was texting with cell phones through email accounts supported by the cell phone providers. The creation of Japanese cell phone email addresses is complicated, because addresses must be long and creative to avoid receiving hundreds of spam messages a day. As a result, the complex and very unique individual email addresses contain many pieces of the owners’ identities. In addition to their first names and birthdays, the participants include some symbols associated with best friends, boyfriends, pets, favorite musicians, and favorite anime characters, foreign languages they are learning, and communities of practice in which they are engaged.

All the participants/pairs were asked the same questions in the same order. The participants seemed comfortable answering the interview questions. For the ratings of femininity and masculinity among the Tokyo, Kyoto, and Osaka varieties, one participant had a problem giving me her judgment. While eliciting the participants’ opinions or beliefs about language, regional, and gender ideologies is more complicated than obtaining information about their demographic background or communities of practice, open-ended questions were effective in bringing out a wide range of responses. To ask questions about images of Japanese Women’s Language (JWL), I first confirmed the participants’ awareness of what is called JWL, by asking, “Have you heard of something called joseego (‘women’s language’) or onna-kotoba (‘women’s language’) in Japanese?” and then asked the questions listed on the questionnaire. All of the participants recognized the terms and the language variety.
The interviews were carried out very smoothly. Most participants used more polite linguistic forms during the interviews than they did during the conversations, but a few continued using the casual register. Interviewing by pairs helped to elicit commentaries from the participants. By having two interviewees, the interactions were not unidirectional, which made the setting more relaxed and conversational. Also, the participants needed to explain their views not only to me but also to their friends, which made them wordy and elaborate. Oftentimes, I received interesting information about a participant from her friend rather than from the participant herself.
CHAPTER 3
Breathy Voice: The Essence of Elegance

3.1. Introduction

This chapter examines breathy voice and regional variation across the Tokyo, Kyoto, and Osaka regions of Japan. The analyses of spectral measurements show that the Osaka participants overall use more breathy voice than the Kyoto participants but reveal that the use of breathy voice in Kyoto is different from the use in Tokyo and Osaka. The Kyoto participants’ use of breathy voice displays a stronger correlation with features stereotypically associated with soft and feminine speech such as low intensity and high pitch than that of the Tokyo and Osaka’s participants, which shows that the Kyoto speakers construct femininities differently from the Tokyo and Osaka speakers. In this chapter, I will argue that breathy voice connotes social meaning only in association with other linguistic features and does not carry meaning alone. Also, this chapter will show that linguistic features that are commonly associated with gender, not region, like breathy voice can exhibit regional variation.

To facilitate the study, the first half of this chapter will be dedicated to discussing the acoustic properties of breathy voice by surveying definitions and measurements of breathy voice. The application of acoustic measurements in sociophonetic studies will be highlighted in the following literature review section focusing specifically on breathy voice and gender. Building on the previous literature, the latter half of the chapter will discuss the present study.

3.2. Background I: Phonetics of Breathy Voice

3.2.1. Phonation
Phonation refers to the manner in which vocal folds vibrate and transform an airstream into sounds. Phonation, with a few exceptional languages, involves a pulmonic egressive airflow across the larynx, which lies between the subglottal and supraglottal cavities. In the larynx, the vocal folds (see Figure 3.1) vibrate with the repetitive actions of open and closed glottal states, which is caused by aerodynamic and myoelastic movements. Therefore, the combination of vocal-fold vibration, frication noise, and the opening and closing of the glottis form the main sources of variation in phonation types or voice qualities such as modal, creaky, and breathy voice.

![Midsagittal section of the vocal tract](image)

**Figure 3.1.** Midsagittal section of the vocal tract (Kent and Read 2002: 9)

By definition, voice quality is realized in vowels and sonorants, since vowels and sonorant consonants require vocal-fold vibration. Different modes of vocal-fold vibration lead to different voice qualities. The states of glottis of different voice qualities are illustrated in Figure 3.2, taken from Catford (1988: 54). When speaking with breathy voice (see the diagram (a) in Figure 3.2),
muscle tension is low, which prevents the vocal cords from closing during vibrating. As a result, the vocal folds do not touch in the middle portion. Therefore, the airstream escapes from the vocal folds and results in air turbulence and audible friction.

Figure 3.2. States of the glottis according to phonatory settings; note that the configuration for breathy voice is the same as that for voicelessness (Catford 1988: 54)

3.2.1.1. Breathy Voice vs. Whisperty Voice

As Laver (1980:133) notes, breathy voice and whisperty voice share the characteristic of audible friction, but the major difference between the two is the level of muscle tension during phonation. While breathy voice is accompanied by low muscle tension, in whisperty voice, the vocal muscles are tense. Laver further points out a possible confusion of the terms as it is a common practice in acoustic research to label a voice quality with any friction components as breathy voice regardless of the degree of the vocal tension. Hammerberg et al. (1986) emphasizes that there are two types of breathy voice: a breathy and hypofunctional (lax) voice and a breathy and hyperfunctional (tense) voice. The former indicates what Laver defines as breathy voice, and the latter is the quality defined as whisperty. According to Hammerberg et al., hyperfunction, coming from compressed vocal folds, results in strained voice, and it is more likely to be found in disordered speech. The present study, while being aware of the possible
confusion, will use the term breathy voice for voice quality with any audible friction because the target participants in this study are non-disordered and because breathy voice has been more often discussed in the context of gender than whispery voice.\textsuperscript{16}

3.2.2. Acoustic Correlates of Breathy Voice

Due to the multidimensional nature of articulatory configurations of breathy voice as shown by Keating et al. (2010), there are a number of acoustic and aerodynamic properties that potentially reflect the degrees of breathiness of voice. In this section, I will focus on the acoustic measures for breathiness since acoustic measures are most suited for my study. Hirano (1981: 66) provides overviews of different methodological approaches used in clinical studies and concludes that acoustic analysis of the voice signal is “the most attractive method” since “it is non-invasive and provides objective and quantitative data.” Most aerodynamic approaches require equipment attached to a speaker’s articulatory organs such as the Rothenberg mask to measure airflow (Rothenberg 1973), a transducer to measure glottal airflow (Cranen and Boves 1985), and the endoscopic and fiberscopic systems to observe the vocal folds. These methods, by restricting free and natural movement of articulatory organs, are not very suitable for natural spontaneous speech, especially conversational data. Conversational data is most appropriate for my study to obtain the participants’ most relaxed speech possible in investigations of their daily linguistic practices.

3.2.2.1. Spectral Measures

\textsuperscript{16} In some languages, it is documented that breathy voice and whispery voice carry distinctive meanings. For example in Lachixio Zapotec, breathy voice connotes authority while whispery voice marks urgency (Sicoli 2010, forthcoming).
Breathiness of voice can be measured based on its spectral properties, which can be observed in a Fast Fourier Transform (FFT) spectrum as presented in Figure 3.3.

![FFT spectra for modal voice (left) and breathy voice (right) of the vowel [a] taken from the Tokyo participant Jun](image)

Figure 3.3: FFT spectra for modal voice (left) and breathy voice (right) of the vowel [a] taken from the Tokyo participant Jun

The difference between the increased amplitude of the first harmonic in a vowel and the amplitude of the second harmonic, \( H_1-H_2 \), correspond to the open quotient of the glottal waveform. Open quotient is defined as a ratio of the time period in which the glottal airflow increases and decreases to the total period of vocal fold oscillation (Hanson 1997), and a larger open quotient is a salient acoustic property of breathy voice. Fisher-Jørgensen’s (1967) perception study shows that \( H_1 \) of Gujarati stimuli more successfully offers the cue to breathy vowels for Gujarati listeners than other factors like duration and \( F_0 \). The correlation between an \( H_1-H_2 \) value and open quotient is also supported in a variety of languages in addition to American English (Holmberg et al. 1995), such as !Xöö (Ladefoged and Antoñanzas-Barros

Some studies use the corrected acoustic measure of open quotient, represented as $H_1^*-H_2^*$, on the basis of the claim that the spectral slope correlates with the abruptness or gradualness of vocal fold closure (Stevens 1977, Stevens and Hanson 1995, Hanson 1997). Corrections are made “for the amounts by which $H_1$ and $H_2$ are 'boosted’ by the effect of the first formant on the vocal-tract transfer function” by subtracting $20\log_{10}(\frac{F_1^2}{(F_1^2-f^2)})$ from $H_1$ and $H_2$, where $f$ is the frequency at which the harmonic is located (Hanson 1997: 475, 480). The effect of this correction is that the amplitudes of the harmonics more closely approximate those of the actual source spectrum, and making this correction is particularly important in comparisons across vowels, for which $F_1$ can vary greatly (Hanson and Chang 1999: 1066). However, as is pointed out by Iseli and Alwan (2004), Hanson’s model assumes that fundamental frequency is at least 100 Hz away from $F_1$ and does not take into account the bandwidth of $F_1$ or the effects of higher formants. Therefore, $H_1^*-H_2^*$ would be useful only if the fundamental frequency is relatively the same across the speakers to be compared.

While the most widely used acoustic measure of phonation is $H_1-H_2$, breathy voice is also associated with a steeper spectral tilt. Spectral tilt is measured as the amplitude difference between the amplitude of the first harmonic and that of one of the first three formants ($A_1$, $A_2$, and $A_3$). $H_1-A_1$, while reflecting the first formant bandwidth ($B_1$), corresponds to the effect of a posterior glottal opening (Garellek and Keating 2010, Hanson et al. 2001), with a higher $H_1-A_1$ value indicating that the vowel is produced with a glottal configuration including a posterior glottal opening. On the other hand, $H_1-A_2$ and $H_1-A_3$ correspond to abruptness of vocal closure. A steeper spectral tilt is caused by less abrupt glottal closure involved in breathy phonation which amplifies $F_0$ and simultaneously dampens the higher harmonics. The reliability of spectral parameters such as $H_1-H_2$, $H_1-A_1$, and $A_1-A_3$ to measure glottal configurations is
confirmed by a comparative study between aerodynamic and acoustic measurements of the female voice conducted by Holmberg et al. (1995).

Although these different measures of a spectral tilt have been used widely in studies to distinguish breathy voice quality from other phonation types, they do not always behave uniformly. Esposito (2007, 2010) summarizes results of previous studies on 10 languages/dialects (Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang, and !Xóõ) where breathy vowels have phonemic contrast with modal vowels and examines the success of the different measures in distinguishing the phonemic breathy vowels from modal vowels. She concludes that the four spectral measures can be ranked by cross-linguistic success in the order of H1-A3, H1-H2, H1-A2, and H1-A1. Although these measures are identified as more or less successful, none of these spectral measures, even the most successful measure H1-A3, distinguish the two phonations of breathy vowels and non-breathy vowels successfully in all the languages, which suggests that multiple measures should be consulted when using these spectral parameters.

3.2.2.2. Periodicity

In addition to the spectral properties, breathy voice is also associated with an increase in aperiodic noise, especially at higher frequencies. This noise indicates substantial noise during breathy vowels due to the persistent air leakage from the glottis. Measures of the noise and periodicity used in literature include cepstral peak prominence (CPP) – a measure of cepstral peak amplitude normalized for overall amplitude (Hillenbrand et al. 1994) – and harmonic-to-noise ratio (HNR) – a measure of harmonic energy normalized by the spectral noise level. Because of the additive noise from the increased airflow, values from all of these measures are lower in breathy voice.
These measures have been used in studies on phonation to distinguish breathy from non-breathy voice qualities. By summarizing literature on 10 different languages/dialects where breathy vowels are phonemically contrastive from modal vowels, Esposito (2007) concludes that CPP is overall a good measure by measuring a difference in noise to distinguish the two phonations. Esposito (2007, 2010) further concludes that CPP is the best of the seven measures (CPP, H1-H2, H1-A2, H1-H3, A2-A3, the average of H1-H2 compared to A1) for determining breathy voice in Jalapa Mazatec, and Blankenship (2002) has also reported that CPP, H1-H2, and H1-A2 are equally effective in distinguishing breathy from modal phonations in Krathing Chong. In Miller’s (2007) study of Ju|’hoansi, HNR, with its lower values, correctly captures the breathiness of gutturals, and Wayland, Gargash and Jongman (1994) report that Javanese clear vowels exhibit higher HNRs than breathy vowels. On the other hand, Wayland’s (1997) study of Chanthaburi Khmer presents a reverse pattern: the HNR values were higher for breathy vowels than those for clear vowels. Ladefoged and Antoñanzas-Barros (1985) argue that spectral tilt is a better measure than the measurements of periodicity based on their finding that there is a considerably lower correlation between the aperiodic cycles in the waveform and American listeners’ perception of breathiness of !Xóõ.

3.3. Background II: Breathy Voice and Gender

3.3.1. Male-Female Differences

With the various research methodologies and measurements, it is consistently reported that a greater prevalence of breathy voice is observed among women than men. This pattern was found cross-linguistically, for American English (Rothenberg 1973, Klatt and Klatt 1990, Hanson and Chuang 1999), British English (Henton and Bladon 1985), Canadian English (Heffernan 2004), Dutch (Günzburger 1991), Spanish (Trittin and De Santos 1995; Mendoza et al. 1996), and
Japanese (Heffernan 2004). The tendency for women to be more breathy than men is often discussed as having physiological roots. Together with earlier clinical studies, Södersten and Lindestad (1990) suggest that 1) females have more incomplete closure than males; 2) the degree of breathiness is perceived to be higher for female voices than for male voices; and 3) it is a common tendency across gender categories that the degree of incomplete closure corresponds to the degree of breathiness. Even though they found the female voice to be more breathy, it was argued that the size of the posterior glottal gap, not the gender category, results in the correlation between perceived breathiness and the degree of incomplete glottal closure. In their study using aerodynamic measurements for the degree of breathiness, Koike and Hirano (1986) hypothesized that the gender differences in mean airflow rate – measurement of the air consumption per unit of time during speech production – arise simply from the different vital capacities between males and females. To minimize the effects of the physiological gender differences, Koike and Hirano proposed a new measure called the Vocal Velocity Index (VVI), which normalizes the mean airflow rate by the subject’s vital capacity. Holmberg et al. (1988, 1989) found that female speakers tend to have larger open quotients and more gradual rises and falls in glottal flow than male speakers, and Klatt and Klatt (1990) and Hanson (1997) show that careful control of glottal characteristics improves the naturalness of synthesized female speech.

Here, we should acknowledge the caution raised by Simpson (2012) regarding using H1-H2 to examine the gender-based difference in breathy voice. This is because the nasal resonance amplifies different harmonics for women and men, even for mostly oral vowels. The nasal resonance shifts H1-H2 higher for women. Therefore, the use of H1 and H2 may find female voices to be breathier than what they actually are, simply due to the effect of spontaneous nasality.
3.3.1.1. Non-Physiologically Based Differences

Although some researchers tend to attribute different levels of vocal breathiness to the differences between male and female physiological attributes, studies show that voice quality can be manipulated in gender performances regardless of the speaker’s physique. For example, it is often found that breathy voice is explicitly manipulated in commercially produced female voices. Starr (2010) show that the voice actors in Japanese anime use “sweet voice”, characterized in part by syllables ending in breathy voice, to construct a figure that is mature and feminine. Compared to characters with “non-sweet” voices examined by Starr (2010), sweet voice characters are relatively older women “in positions of traditional female authority,” who are, for example, mothers, older sisters, and goddesses. The association between Japanese anime culture and manipulated breathy voice was also found in Kawahara (2013). Kawahara studied the speech of Japanese “maid” girls who work at theme restaurants dressing in maid costumes. Deriving from popular male-targeted anime, the maid girls’ speech exhibits features commonly associated with the speech of those anime characters. Using the readings of five vowels, Kawahara (2013) noted that when speaking in their maid voice, they use more breathy voice than when speaking in their normal voice. Interestingly, across the speakers, breathy voice was commonly used at the second to the last vowel position regardless of the order of the vowels.

Hall (1995) reported that both male and female phone sex workers in the U.S. strategically use breathy voice in the performance of women’s sexual arousal. Kajino and Moon (2011) also found that Japanese porn actresses use increasingly more breathy voice when talking about “sexual topics” than when talking “non-sexual topics” in order to portray themselves as sexual beings.

Professional actors utilize breathy voice to express certain emotions. Guzman et al. (2013) had both female and male Spanish actors read a passage aloud with six different emotions (happiness, sadness, fear, anger, tenderness, and eroticism) as well as at the neutral state. The finding from the spectral measures showed that both males and females use more
breathy voice when expressing eroticism, sadness, and tenderness than when they express joy and anger, and this finding was also supported by the listeners' ratings on breathiness as well. Surprisingly, Guzman et al. found that male actors' speech was more breathy than female actors, although they noted that using the mean of all the samples from different emotional performances was not a compatible method to other studies on breathy voice and gender.

Manipulation of breathy voice can be observed in real life contexts as well. Ito (2005) argued that young adult Japanese males use more breathy voice when talking to socially superior interlocutors to index politeness. In Lachixio Zapotec as well, as Sicoli (2010, forthcoming) demonstrates, voice quality serves as an indexical cue to the relative status of speakers and addressees in interactions. Breathy voice marks authority, by co-occurring with low pitch. In their study on American male-to-female (MTF) transgender speakers by Gorham-Rowan and Morris (2006), the flow inverse filtering method show significantly higher airflow parameter values in the participants' female voices than in their biological male voices. Holmberg et al.'s (2010) study on 25 MTF transgender speakers presents a similar finding with relatively high average airflow values, even though perceptual ratings of breathiness were low. These two studies clearly show that breathy voice is a feature commonly used by trans women in their attempts to attain female voice, though it should also be noted that neither found that breathiness is significantly related to perceptual gender ratings or the voice that was successfully perceived as female.

3.3.2. Perception Studies

Perceptual ratings of listeners are often used to identify personality traits associated with breathy voice. In Pittam (1987), 6 female and 6 male Australian speakers read a text with different voice qualities, including breathy voice, and the speech material was judged by 80
Australian and 80 American students on perceived status (e.g. profession, ambition, intelligence, dominance) and solidarity (e.g. friendly, sympathetic, likeable, trustworthy). Breathy voice in female speakers scored significantly higher on solidarity than other voice qualities in female speakers as well as breathy voice in male speakers. Also, breathy voice by women was rated as the highest status vs. other phonation types by women, whereas tense voice\textsuperscript{17} was judged as the highest status among the different voice qualities by men. Addington (1968) demonstrated that the same voice feature can be interpreted differently in a male and female voice. In this study, 2 female and 2 male speakers were recorded during a read-aloud task of texts in a breathy voice, and personality traits of the speakers were judged by 320 listeners. The results showed that females simulating increased breathiness were perceived as being more feminine, prettier, more petite, idealistic and shallower, whereas male voices with the same voice quality were associated with youth and an artistic quality.

3.3.3. Age

Age can be observed in the interactions with gender in production of breathy voice. Among English speaking children between ages 4 and 8 in Ferrand’s (2000) study, it was found that male children use more breathy voice than female children. In Chen et al.’s (2010) examinations of children between ages 8 and 17, the male children start exhibiting higher CPP values (i.e. less breathy) than female children at the ages 10-11, and the difference becomes larger as age increases with a continuous rise of male CPP values. The CPP values of female children very gradually increase until ages 14-15, but interestingly, decrease at ages 16-17. Furthermore, Ferrand (2002) reported that elderly females aged 70 to 90 are more breathy than young female

\textsuperscript{17} Tense voice is characterized by a tense vocal tract musculature and accompanied by vigorous movement of the vocal organs (Laver 1980). As noted above, tense voice contrasts with breathy voice, which is accompanied by low muscle tension.
adults aged 21 to 34 and middle-aged females aged 40 to 63. However, it should be noted that the breathy voice in elderly adults in her data co-occurs with a significantly lower pitch level than that of young adults and middle-aged adults. The combination of breathy voice with lower pitch establishes a different social meaning than when a breathy voice occurs with higher pitch. This point will be closely investigated later in this chapter and in Chapter 6 of this dissertation.

3.3.4. Dialects

Finally, dialectal variation in breathy voice is reported across a few language communities. Henton and Bladon (1985) found for British English that both male and female Modified Northern speakers use more breathy voice than the Received Pronunciation (RP) speakers. Purnell, Idsardi, and Baugh (1999) showed that speakers of African American Vernacular English are significantly more breathy than speakers of Chicano English and Standard American English. Variation in breathy voice across White and African Americans is further examined by Podesva (2013). In his work on voice quality features in Washington, D.C., he discovered that whereas females are more breathy than males among White Americans, males are more breathy than females among African Americans. While the pattern among white speakers is compatible with many other studies, incorporating African American populations reveals a new perspective. It is white men that are least breathy, which may suggest that breathy voice indexes (white) femininity on the one hand, and low rates of breathiness may carry the association with white masculinity on the other.

3.4. Current Study

18 In the study, the speakers of Modified Northern British English are defined as people who grew up in the area of Leeds but have moved away for substantial periods and have thus modified their native features (Henton and Bladon 1985: 224).
3.4.1. Introduction

Building on the previous studies of breathy voice and gender reviewed in the previous section, the current study investigates breathiness of voice and its interactions with other acoustic correlates by the participants from Tokyo, Kyoto, and Osaka. The analyses draw on the naturalistic conversations I recorded with the 46 participants at women’s colleges in the three regions. Details of the data, participants, and recording method are presented in Chapter 2. The results indicate that the Osaka participants generally use higher degrees of breathy voice, but I argue that the ways in which breathy voice is used holds more significance in the construction of femininities than overall degrees of breathiness alone since breathy voice occurs through various causes, including physiological ones as we have seen in the previous section (e.g. Södersten and Lindestad 1990). Instead, examination of co-occurring features of breathy voice provides more essential information on the speaker’s gender construction than overall degree of breathy voice that may be caused by the speaker’s physiological characteristics. While the Kyoto participants use breathy voice with soft and feminine features like low intensity and high pitch, the Tokyo and Osaka participants’ use of breathy voice are not linked with those features. The particular feminine image constructed by the differing patterns of the acoustic correlates will be discussed in the discussion section.

3.4.2. Procedure

For the analysis, four spectral parameters of glottal configuration, all of which are associated with perceived breathiness, are measured through a target vowel. As is reviewed in the previous sections (especially, see Hanson 1997, Hanson et al., 1999, 2001), the spectral measurements are theoretically supported. First, a change in an open quotient during breathy phonation affects the glottal waveform to more closely approximate a sinusoid of frequency Fo, and therefore, the
amplitude of the first harmonic (H1) increases relative to the amplitudes of the higher harmonics (H2). The measure H1-H2 has been used to reflect the open quotient in previous research. Second, as the first-formant bandwidth (B1) increases, the amplitude of the first-formant peak (A1) decreases. Therefore, the relative amplitude of the first harmonic and the first-formant peak (H1-A1) is used as an indicator of B1. Third, the spectrum is a derivative of the glottal waveform at middle and high frequencies and is influenced by the abruptness with which airflow is cut off when the membranous part of the vocal folds closes during the vibratory cycle. Therefore, the amplitude of the first harmonic compared with the amplitudes of the second- and third-formant peaks (H1-A2 and H1-A3) is an indicator of spectral tilt at the mid- and high-formant frequencies.

To conduct acoustic analysis, the vowel [a] was selected as the target vowel. Among the five Japanese vowels ([i], [e], [a], [o], [u]), the first and second formants of [a] are most likely to be separable from each other. In order to avoid having the effect of the first formant on H1 and H2, I concentrate on one vowel. The vowels that occur in elongated environments are excluded since there are significant effects of the vowel length in their formants (Akaba 2008). Any tokens where the boundaries with the preceding and following sounds are unclear were also eliminated. Vowels at the IP (intonational phrase) final positions were avoided due to their tendency to be produced with creaky voice (Henton and Bladon 1988; Kawahara and Shinya 2008 on Japanese). Also vowels at the IP initial positions were excluded since my preliminary analysis found that the IP initial vowel [a] tends to be more breathy than non-IP initial [a]. This finding is also supported by Podesva (2013), as the further a vowel is positioned from the beginning of IPs, the less likely it is to be produced with breathy voice because of the difficulty associated with sustaining breathy voice (Catford 1977).\footnote{I did not include phrase positions (the distance – in number of mora – from the initial and final IP boundaries) in the coding since the speakers’ speech rates significantly vary across regions (the}
geminates were excluded based on previous discussion of laryngeal affects of gemination.

According to Kawahara (forthcoming: 22-23), the articulatory nature of Japanese gemination is still to be explored, and whether geminates involve laryngeal constrictions is debatable. Even though Fujimoto et al. (2010; cited in Kawahara forthcoming) did not find any laryngeal constrictions in their experiment using a high-speed digital video recording system, Idemaru and Guion’s (2008) study which reports shallower spectral tilt in H1-A1 in the vowels following geminates suggests some potential effect of geminates in voice quality.

100 to 120 tokens of the vowel [a] were coded for each speaker. Initially, 120 tokens were coded for all the speakers, but the variation in the number of tokens was caused by eliminating all the tokens with measurement failures where no values were recorded. In total, 1,637 tokens were coded for Kyoto, 1,792 tokens were coded for Osaka, and 1,710 tokens were coded for Tokyo.

Using a script provided by the UCLA PRAAT Script Resources as the basis, I prepared a PRAAT script to measure H1-H2, H1-A1, H1-A2, and H1-A2 of each token of [a]. The script goes through the entire tier specified within a text grid and makes those acoustic parameter measurements for each interval of [a]. Each interval of [a] was segmented into three, and measurements were taken only from the segment in the middle in order to minimize affects of co-articulation of preceding and following segments. The parameters measured are summarized for convenience as below.

**H1-H2**: the relative amplitudes of the first and second harmonics, reflecting an open quotient

**H1-A1**: the relative amplitudes of the first harmonic and the first formant, reflecting

Kyoto and Osaka participants exhibit significantly higher speech rates than the Tokyo participants), which interrupts cross-regional comparisons.
the first formant bandwidth

**H1-A2**: the relative amplitudes of the first harmonic and the second formant, reflecting spectral tilt

**H1-A3**: the relative amplitudes of the first harmonic and the third formant, reflecting spectral tilt

Additionally, duration, fundamental frequency (F0), and intensity of [a] were measured to examine internal linguistic constraints of the phonation of [a]. The mean F0 of each vowel segment was measured with F0 values at every 5 ms. Intensity of the vowel was normalized as *relative intensity* by subtracting the intensity of the preceding consonants from the intensity of the vowel [a]. For independent variables, speakers and regions were included. A series of mixed effects linear regressions were conducted, employing acoustic parameters as dependent variables. Duration, relative intensity, F0 average and region were treated as independent variables to examine the interaction of their acoustic correlates with regions in the use of breathy voice. Speaker was included as a random effect.

### 3.4.2.1. Additional Linguistic Features

Before moving onto the next section of results, I will briefly review gender meanings of the additional linguistic features examined here. As variation in duration of the vowel [a] has more to do with speech rates than the gender meanings\(^{20}\), I will discuss only F0 and intensity.

**Fundamental Frequency (F0)**

\(^{20}\) Speech rates are also associated with gender, but the gender meanings do not directly link to the breathiness of [a], so I will postpone the discussion of speech rates until Chapter 6.
Fo levels have been one of the most frequently examined features in connection with gender, and higher Fo levels are commonly associated with female voices. Ohala (1984) studied the meaning of Fo levels in speech cross-linguistically and developed the notion of “frequency code,” where low pitch/high pitch is equivalent to strong/weak or large/small. According to Ohala, the “frequency code” also governs the vocalizations of other species and associates the primary meaning of “small vocalizer” (high Fo) with the secondary meanings such as “subordinate,” “submissive,” “non-threatening,” and “desirous of the receiver’s goodwill,” while associating the meaning of “large vocalizer” (low Fo) with meanings such as “dominant,” “aggressive,” and “threatening” (4-5).

While these connotations may be rooted in physical correlations between vocal tract size and physical size of human/animal bodies, physiology is not the only factor affecting variation in Fo or its social meanings. For example, Van Bezooijen’s (1995) study comparing Dutch and Japanese showed that Japanese women have significantly higher Fo levels than Dutch women and claimed that the difference between Japanese women’s and Dutch women’s Fo levels is related to the different images of “ideal” women in the two societies. While Japanese listeners find the higher pitch of female voices more attractive, Dutch listeners perceive the mid and lower pitches of female voices more attractive. Yuasa (2011) showed that Japanese women also make use of a higher mean pitch than American women. In Ohara’s (1999, 2001) studies of American and Japanese bilingual speakers of Japanese and English, it was found that both Japanese and American female speakers employed a higher Fo when speaking Japanese than when speaking English. In the interviews, these American English-Japanese bilingual women expressed their awareness that a higher voice pitch is one of the key components for establishing femininities in Japanese society (Ohara 2001). Thus, higher Fo levels in Japanese carry a particularly powerful ideological link to traditional womanhood since it is considered as a feature of JWL (Horii 1990, Ide 1990).
Voice pitch in Japanese is also investigated in terms of more specific personality traits. Teshigawara (2003) examined phonetic properties of heroes’ and villains’ voices in Japanese animes (‘animations’) and found that mean Fo does not vary between the two. Van Bezooijen’s (1996) perception study found that a higher pitch evokes images such as being dependent, modest, less prestigious, more sensitive, and emotional, while a lower pitch is associated with qualities such as being independent, arrogant, prestigious, less sensitive, and more rational. Certain personality traits such as dependence, subservience, modesty, innocence, and helplessness are traditionally more valued for women in Japanese culture than in North American culture (Smith 1992), and it is certainly not a coincidence that those are the traits invoked by a higher pitch.

Voice pitch in Japanese is further associated with speakers’ sexual orientation. Camp (2009) examined the voice pitch of Japanese lesbian and heterosexual speakers. The perception experiment of the study showed that lower Fo and lower Fo range significantly correlated with listeners’ judgment that the speaker is a lesbian. At the same time, these features judged to be lesbian were also judged to be more masculine and less emotional than those judged to be heterosexual. Analysis of speech production supported listeners’ perceptions of lesbian speech, as the mean pitch produced by lesbian speakers was significantly lower than that of heterosexual speakers. Camp’s study shows that a type of femininity connoted by a higher pitch is a \textit{heterosexual} femininity.

\textbf{Intensity}

Compared to Fo, intensity and its association with gender are not extensively studied in sociolinguistics. One exceptional study, Aronovitch (1976) investigates correlations between mean intensity levels of speech in American English and listener’s evaluations of speakers’
personality traits. For female voices, speech with lower intensity was rated as more self-doubting, introverted, cautious, and submissive, while speech with higher intensity was rated as more self-confident, extroverted, bold, and dominant. However for male voices, no correlations were found between the listeners’ evaluations and varied mean intensity levels, which indicates that overall intensity levels are a significant factor for people’s impressions about women but not about men. Conversely, in their study with 30 male psychiatric patients hospitalized at the Veterans Administration Center in Jackson, Mississippi, Eisler et al. (1973) found that loudness of speech positively correlated with the levels of perceived assertiveness. In either case, the attributes associated with lower intensity (such as self-doubt, introversion, caution, and submissiveness) in female voices correspond with what is considered as a traditional feminine figure in Japan more than the attributes associated with higher intensity levels among women (such as self-confidence, extroversion, boldness, dominance, and assertiveness).

3.4.3. Results

Results of the regression analysis conducted with R are summarized in Tables 3.1-3.4. The intercept indicates the estimated values of the spectral measures (i.e. H1-H2, H1-A1, H1-A2, and H1-A3) if all independent variables (F0, duration, relative intensity, and region) are treated as the intercept. In these particular models, Kyoto is chosen as the intercept. To determine estimated values for the spectral measures given non-default values for the independent variables, each independent variable’s coefficient is added to the intercept’s estimated coefficient.

---

21 In clinical research, Brockmann et al. (2008) examined voice loudness and gender effects on measured sound pressure level, F0, jitter, and shimmer in the vowel /a/ at three different loudness levels: soft, medium, and loud. The results showed that in soft and medium phonation, men were generally louder with higher SPL values, which may suggest a connection between loudness and male speech.
Tables 3.1-3.4 summarize the results of the regression models, and Figure 3.4 shows cross-regional variation in the mean values and distribution of the spectral measurements. The mixed effects model finds Fo and intensity to be significant predictors for all the four parameters while duration is found to be significant only in H1-A2. As is seen from the positive estimate for Fo, the Fo levels positively correlate with the values of H1-H2, H1-A1, H1-A2, and H1-A3, which indicates that there is a tendency for the breathiness of the vowel [a] to increase as the Fo levels increase. On the other hand, the intensity levels, with the negative estimate, present negative correlations with those metrical values. This means that the vowel [a] is less likely to be breathy when the intensity is higher. These two factors, Fo and intensity, will be examined in more detail later in this section. As for the factor duration, its negative coefficient for H1-A2 indicates a negative correlation between duration of the vowel [a] and the H1-A2 values. Higher H1-A2 values are likely to be found when the vowel duration is shorter. H1-A2 corresponds to abruptness of vocal closure, with a higher value being caused by less abrupt glottal closure involved in breathy phonation. The pattern shows that the vowel [a] is likely to be more breathy when its duration is shorter.

| Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|----------|----------|------------|------------|-------|---------|
| (Intercept) | -15.1406 | -15.1952 | -19.4748 | -10.7401 | 0.0001   | 0.0000   |
| Duration | -0.0212 | -0.0210 | -0.0486 | 0.0047 | 0.1182   | 0.1209   |
| F0.ave | 0.0726 | 0.0728 | 0.0572 | 0.0893 | 0.0001   | 0.0000** |
| Intensity | -0.4375 | -0.4369 | -0.5662 | -0.3006 | 0.0001   | 0.0000** |
| Regionosaka | 5.5392 | 5.6352 | -0.4528 | 11.5574 | 0.0690   | 0.0768   |
| Regiontokyo | 4.9725 | 5.0748 | -0.8003 | 11.5625 | 0.1056   | 0.1129   |
| Regionosaka: Duration | 0.0324 | 0.0320 | -0.0022 | 0.0697 | 0.0774   | 0.0815   |
| Regiontokyo: Duration | 0.0236 | 0.0231 | -0.0137 | 0.0563 | 0.1930   | 0.1883   |
| Regionosaka: Intensity | 0.2367 | 0.2360 | 0.0445 | 0.4104 | 0.0138   | 0.0108*  |
| Regiontokyo: Intensity | 0.2252 | 0.2236 | 0.0425 | 0.3899 | 0.0098   | 0.0107*  |
| Regionosaka: F0.ave | -0.0287 | -0.0290 | -0.0514 | -0.0070 | 0.0110   | 0.0109*  |
| Regiontokyo: F0.ave | -0.0365 | -0.0368 | -0.0593 | -0.0145 | 0.0024   | 0.0012** |

Table 3.1. Mixed effects for model of H1-H2
| Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|----------|----------|------------|------------|-------|----------|
| (Intercept) | -22.6369 | -22.6505 | -27.4077 | -17.7381 | 0.0001 | 0.0000 |
| Duration | -0.0063 | -0.0063 | -0.0363 | 0.0226 | 0.6828 | 0.6807 |
| F0.ave | 0.0901 | 0.0901 | 0.0726 | 0.1084 | 0.0001 | 0.0000** |
| Intensity | -0.5202 | -0.5195 | -0.6799 | -0.3747 | 0.0001 | 0.0000** |
| Regionosaka | 10.2778 | 10.2566 | 3.8325 | 17.0266 | 0.0028 | 0.0025** |
| Regiotokyo | 4.7314 | 4.7508 | -2.0450 | 11.3619 | 0.1658 | 0.1651 |
| Regionosaka: Duration | 0.0207 | 0.0208 | -0.0212 | 0.0602 | 0.3188 | 0.3206 |
| Regiontokyo: Duration | 0.0171 | 0.0170 | -0.0233 | 0.0552 | 0.3986 | 0.3928 |
| Regionosaka: Intensity | 0.1791 | 0.1770 | -0.0236 | 0.3881 | 0.0942 | 0.0850 |
| Regiontokyo: Intensity | 0.1691 | 0.1681 | -0.0221 | 0.3639 | 0.0882 | 0.0866 |
| Regionosaka: F0.ave | -0.0354 | -0.0352 | -0.0602 | -0.0111 | 0.0046 | 0.0050** |
| Regiontokyo: F0.ave | -0.0229 | -0.0228 | -0.0469 | 0.0025 | 0.0710 | 0.0702 |

Table 3.2. Mixed effects for model of H1-A1

| Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|----------|----------|------------|------------|-------|----------|
| (Intercept) | -6.9411 | -6.9569 | -12.0516 | -2.0108 | 0.0054 | 0.0084 |
| Duration | -0.0545 | -0.0545 | -0.0850 | -0.0248 | 0.0008 | 0.0003** |
| F0.ave | 0.0752 | 0.0752 | 0.0577 | 0.0931 | 0.0001 | 0.0000** |
| Intensity | -0.4032 | -0.4020 | -0.5505 | -0.2535 | 0.0001 | 0.0000** |
| Regionosaka | 10.6356 | 10.6565 | 3.6515 | 17.6655 | 0.0034 | 0.0034** |
| Regiotokyo | 6.3783 | 6.4577 | -0.7243 | 13.2406 | 0.0718 | 0.0800 |
| Regionosaka: Duration | 0.0187 | 0.0187 | -0.0210 | 0.0612 | 0.3676 | 0.3666 |
| Regiontokyo: Duration | 0.0060 | 0.0059 | -0.0341 | 0.0455 | 0.7754 | 0.7649 |
| Regionosaka: Intensity | 0.1423 | 0.1409 | -0.0645 | 0.3385 | 0.1732 | 0.1680 |
| Regiontokyo: Intensity | 0.1776 | 0.1756 | -0.0097 | 0.3766 | 0.0738 | 0.0699 |
| Regionosaka: F0.ave | -0.0339 | -0.0338 | -0.0579 | -0.0087 | 0.0076 | 0.0070** |
| Regiontokyo: F0.ave | -0.0176 | -0.0178 | -0.0424 | 0.0070 | 0.1592 | 0.1618 |

Table 3.3. Mixed effects for model of H1-A2
|                    | Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|--------------------|----------|----------|------------|------------|-------|---------|
| (Intercept)        | 1.3702   | 1.3476   | -5.3739    | 8.2719     | 0.7000| 0.7073  |
| Duration           | -0.0128  | -0.0120  | -0.0490    | 0.0258     | 0.5248| 0.5033  |
| F0.ave             | 0.0837   | 0.0836   | 0.0621     | 0.1060     | 0.0001| 0.0000**|
| Intensity          | -0.4348  | -0.4325  | -0.6204    | -0.2532    | 0.0001| 0.0000**|
| Regionosaka        | 9.6088   | 9.6155   | 0.4539     | 19.1883    | 0.0486| 0.0568  |
| Regionkoyo         | 9.8036   | 9.8588   | 0.4522     | 19.2479    | 0.0412| 0.0532  |
| Regionosaka:       |          |          |            |            |       |         |
| Duration           | 0.0082   | 0.0071   | -0.0433    | 0.0583     | 0.7810| 0.7527  |
| Regionkoyo:        |          |          |            |            |       |         |
| Duration           | -0.0199  | -0.0209  | -0.0698    | 0.0279     | 0.4098| 0.4281  |
| Regionosaka:       |          |          |            |            |       |         |
| Intensity          | 0.1914   | 0.1896   | -0.0588    | 0.4444     | 0.1402| 0.1408  |
| Regionkoyo:        |          |          |            |            |       |         |
| Intensity          | 0.0991   | 0.0959   | -0.1410    | 0.3388     | 0.4348| 0.4217  |
| Regionosaka:       |          |          |            |            |       |         |
| F0.ave             | -0.0311  | -0.0310  | -0.0613    | -0.0008    | 0.0474| 0.0493* |
| Regionkoyo:        |          |          |            |            |       |         |
| F0.ave             | -0.0232  | -0.0230  | -0.0544    | 0.0081     | 0.1538| 0.1436  |

Table 3.4. Mixed effects for model of H1-A3
Region was found to be a significant predictor of H1-A1 and H1-A2. As the positive coefficient for Region Osaka in Tables 3.2 and 3.3 indicates, Osaka exhibits significantly higher H1-A1 and H1-A2 values than Kyoto. These two measures show that the Osaka participants generally use more breathy voice than the Kyoto participants.
The interactions between region and intensity or F0 present interesting patterns. The interaction between H1-H2 and the intensity levels is significantly different between Kyoto and the other two regions. As is clear from Figure 3.5, Kyoto exhibits a steeper tilt of the trend line compared to Tokyo and Osaka. The H1-H2 values of Kyoto negatively correlate with the intensity levels. This reveals that the Kyoto speakers use more breathy voice when intensity is
lower, while degrees of breathiness of the Tokyo and Osaka speakers’ voices are consistent regardless of the intensity levels.

There is also a significant difference in the interaction between Osaka and Fo for all four metrics, while the interaction between Tokyo and Fo is significantly different from that of Kyoto only in H1-H2. Figure 3.6 exhibits the contrasts in trend lines between Kyoto and the other two regions.
Compared to the trend lines in the Tokyo and Osaka charts, the trend line in the Kyoto figure indicates a stronger correlation between F0 levels and H1-H2 values. Higher H1-H2 values are indicative of a larger open quotient, which is a salient acoustic property of breathy voice. Therefore, the Kyoto speakers use more breathy voice at higher F0 levels, whereas the Tokyo and Osaka speakers’ use of breathy voice does not relate to F0 levels as strongly.

3.4.4. Discussion

The overall pattern shows contrasts between Kyoto and the other two regions, and particularly, Kyoto differs from Osaka the most. Considering that breathy voice is often associated with feminine figures in previous literature (Addington 1968, Pittam 1987, Starr 2010), it is surprising that the Osaka speakers generally use more breathy voice than the Kyoto speakers. Recall that the Osaka language variety is considered as “unfeminine” compared to the Kyoto and Tokyo varieties as discussed in Chapter 1. However, I argue that it is not breathy voice alone that carries the particular social meaning since breathy voice occurs through various causes. For instance, a speaker’s high degree of breathiness may be caused by the small posterior glottal gap rather than gender categories or projection of gender identity. Therefore, examination of how breathy voice co-occurs with other features is more telling than the levels of breathiness. Furthermore, the social meaning emerges only in connection with other features, as the Kyoto patterns show.

The observed pattern of breathy voice at lower intensity in the Kyoto speech is not surprising, since breathy voice is typically associated with a reduction in acoustic intensity (Fischer-Jørgensen 1967 on Gujarati, Thongkum 1988 on Kui and Chong, and Traill and Jackson 1987 on Tsonga). The pattern found in Kyoto is in line with the previous acoustic research, while the patterns in Tokyo and Osaka go against previous findings. As is discussed
earlier, a decrease in intensity connotes attributes that are associated with traditional femininity such as self-doubt, introversion, caution, and submissiveness. This means that the Kyoto participants use breathy voice along with a feminine feature, which resonates with the previously proposed indexical meanings of breathy voice (e.g. mature femininity (Starr 2010), a pretty and petite female figure (Addington 1968), upper-class femininity (Pittam 1987)) as well as the classy image of the Kyoto variety presented in Chapter 1.

What is particularly interesting in the pattern of Kyoto is the co-occurrence of breathy voice and high Fo since breathy voice has been noted to co-occur with lower Fo (Hombert, Ohala, and Ewan 1979, Laver 1991: 203, Sicoli 2010: 523). In the discussion of breathy voiced consonants in various languages such as Punjabi, Tibeto-Burman, and Hindi, Hombert, Ohala, and Ewan (1979: 47-48) explain the possible contributing factors of breathy voice in lowering Fo while noting that more investigation is needed. During phonation, the vocal cords are not closely adducted, and therefore, the air pressure is weak. The subglottal pressure is also markedly lower upon the release of the breathy voiced consonants because of the high rate of airflow, which by itself leads to a lower Fo. Additionally, breathy voice involves less forceful contraction of the laryngeal adductor muscles, which not only act to bring the vocal cords together but also lower Fo. In spite of this theoretical support on the basis of the articulatory configurations, in the present study, breathy voice is used with higher pitch by the Kyoto speakers. This suggests that use of the two features together should be a consequence of a socially driven motivation rather than a physiological cause.

On the other hand, breathiness of the Osaka speakers’ and, to a lesser extent, the Tokyo speakers’ voices is not linked to the Fo across the different metrics. While soft and feminine linguistic features such as high pitch and low intensity accompany the breathy voice of Kyoto speakers, the use of breathy voice by the Tokyo and Osaka speakers does not carry an association with those features, which shows different ways of constructing femininities between
Kyoto and the other two regions. I argue that breathy voice works as a linguistic resource to attain the traditional feminine images by the Kyoto participants, but it is not the case for the Tokyo and Osaka participants. Usage of breathy voice found only in Tokyo and Osaka will be further investigated in Chapter 6.

Finally, while previous sociolinguistic studies of breathy voice were mainly concerned with aspects of gender construction, this chapter has shown that breathy voice can also exhibit regional patterns in addition to gender. This suggests that researchers of dialectal variation should consider features commonly believed to be relevant to speakers’ gender as well as region, for a more comprehensive understanding of regionality.

3.5. Conclusion

This chapter has investigated breathiness of voice in the speech of participants from Tokyo, Kyoto, and Osaka. While results show that, overall, Osaka participants use more breathy voice than the Kyoto participants, I have argued that breathy voice alone does not carry social meaning since there are various factors that cause breathy voice including physiological ones. Instead, breathy voice gains meaning only in connection with other features, so it is important to investigate the ways in which breathy voice is used. Breathy voice is used with other traditional feminine features such as lower intensity and higher pitch in Kyoto, but breathy voice is used regardless of the intensity levels and pitch in Osaka. The pattern of Tokyo is more similar to Osaka than to Kyoto, but the contrast between Kyoto and Osaka is more prominent since Kyoto and Osaka display significant differences in more measurements. These clear contrasts between Kyoto and Osaka are surprising given that the two regions have been described as part of the same dialect area in Japanese Dialectology.
The use of breathy voice with high pitch found in Kyoto is particularly salient in construction of femininities since the co-occurrences of the two features are typically constrained by articulatory properties. Therefore, this pattern is strong evidence of a socially acquired gender behavior in the use of breathy voice among the Kyoto speakers. Meanwhile, strong associations between breathy voice and other stereotypically “feminine” features are not observed in Tokyo and Osaka, which indicates that the Tokyo and Osaka speakers construct femininities differently from the Kyoto speakers. In Chapter 6, meanings of breathy voice particularly found in Tokyo and Osaka will be further explored.

This chapter has argued that interactions of multiple linguistic features, in addition to the patterning of one independent variable, should be considered for the analysis of regional varieties. Also, I have demonstrated that regional variation can be observed in the linguistic features that are commonly considered to be more relevant to gender than to region.
CHAPTER 4

Fricatives: The Source of Noise

4.1. Introduction

This chapter examines acoustic characteristics of voiceless sibilant fricatives /s/ (e.g. sumi ‘charcoal’, kasa ‘umbrella’) and /ɕ/ (e.g. shashin ‘picture’, shumi ‘hobby’) in the speech of the 46 participants from Tokyo, Kyoto, and Osaka. While Chapter 3 demonstrated that a linguistic feature that is commonly linked to gender can show regional variation, this chapter investigates how variants that serve as regional markers simultaneously connote meanings that can be relevant to gender. While realizations of the fricatives index the regionality of Tokyo and Osaka, the features index images of sharpness and clarity. These sound images contribute to the construction of gendered figures that typify each region.

In Chapter 1, I described the distinct regional stereotypes of the three language varieties based on the interview data I obtained from the participants. According to the participants’ metalinguistic commentaries, the Tokyo variety is considered a boring standard variety, while the Kyoto variety has the classy geisha image. The Osaka variety is associated with the harsh and vulgar auntie figure. As a reminder, the word clouds of the regional stereotypes presented in Chapter 1 are repeated below. With these regional stereotypes in mind, the present chapter showcases how realizations of the fricatives correspond with these stereotypes with their regional and non-regional meanings. Non-regional meanings are drawn from the psychoacoustic notions of sound images.
I begin this chapter by highlighting the articulatory configurations of the fricatives in questions and different acoustic measurements of these fricatives. Then, after previous literature on gender and fricatives is reviewed, I will outline psychoacoustic studies and some relevant measurements of sound images. The remainder of this chapter will concentrate on the procedure, results, and discussion of the current study.

4.2. Background I: Phonetics of Sibilant Fricatives

4.2.1. Voiceless Sibilant Fricatives
The voiceless sibilant fricatives are characterized by high intensity, aperiodic sound energy. The sound energy is produced by airflow from the lungs passing through a narrow constriction created in the vocal tract. This constriction is caused by a stricture of articulators without making a complete, airtight closure of the vocal tract. As a result, air is forced out of the mouth through the constricted passageway. When airflow passes along the articulators at a relatively high pressure, turbulence or friction is generated. For instance, as illustrated in Figure 4.1, in articulation of /s/, the tongue tip approximates the alveolar ridge while leaving narrow path for airflow to continue which causes turbulence. Contrastingly, /ʃ/ is made further back than /s/, with the blade of the tongue making the constriction. Thus, the nature of noise depends on which articulators are used and how they are positioned. In addition, as is discussed later, the size and shape of the oral cavity is also a significant factor to characterize fricative noise.

Figure 4.1. Midsagittal sections during the production of the two fricatives (Stevens 1998: 380); the following vowel contexts are /a/ (solid line) and /i/ (dashed line) for /s/ and /a/ (solid line) and /u/ (dashed line) for /ʃ/
4.2.2. English Fricative vs. Japanese Fricative Articulation

Just like English, Japanese has two contrastive voiceless sibilant fricatives. However, the articulatory properties as well as the phonological status of the two fricatives are different in Japanese and English. While the English /s/ is apico-alveolar, Japanese /s/ is more of a laminal-dental sound (Akamatsu 1997). Also, the Japanese /s/ has been shown to be less intense and less sibilant than the English /s/ (Akamatsu 1997). For the counterpart of /s/, the English posterior fricative is coronal post-alveolar /ʃ/, while the Japanese posterior fricative is alveolopalatal /ɕ/. While previous studies have demonstrated that English /s/ and /ʃ/ differ robustly in front cavity size, the Japanese fricatives /s/ and /ɕ/ do not differ in front cavity size as much. In English, /ʃ/ opens up a sublingual cavity by the apical post-alveolar posture with lip rounding to further increase the anterior resonant cavity, which enhances the contrast with /s/. By contrast, the tongue tip is not raised in the production of /ɕ/ and the lips are spread rather than protruded. (Toda and Honda 2003, Narayaman, Alwan, and Haker 1995, Ladefoged and Maddieson 1986, Holliday, Beckman, and Mays 2010, Li et al. 2011, Li, Edwards, and Beckman 2007)

The primary contrast between the English /s/ and /ʃ/ is their tongue positions and the resulting front cavity size, whereas tongue posture is the primary contrast between the Japanese /s/ and /ɕ/, with /s/ having a laminal or dental constriction, and /ɕ/ having the longer “palatalized” constriction (Toda and Honda 2003, Holliday, Beckman, and Mays 2010, Li et al. 2011, Li, Edwards, and Beckman 2007). Due to the approximation of the places of articulation of Japanese /s/ and /ɕ/, the phonological status of the two is controversial. Conventional accounts of the Japanese phonological system claim that a surface [ɕ] alternates with /s/ when preceding /i/ and /e/ (e.g. Hattori 1960, Okada 1991, Ito and Mester 2006, Toda 2007). However, more recent studies (e.g. Wanrooij 2007, Holliday 2007, Li et al. 2007, Holliday et al. 2010, Li et al. 2011) treat them as contrastive phonemes rather than mere allophones. Wanrooij (2007) points
out that /s/ and /ɕ/ make minimal pairs such as *sakai* ‘boundary’ vs. *shakai* ‘society’ and *sumi* ‘charcoal’ vs. *shumi* ‘hobby.’ While acknowledging these conflicting views, I treat /s/ and /ɕ/ as phonemic to make the present study compatible with other cross-linguistic studies of fricatives.

Before moving to the next section, it is important to reemphasize that when distinguishing the two voiceless fricatives, the places of articulation of Japanese fricatives do not carry information as critical as those places do in English. Holliday et al. (2010) found that the distinction between the two Japanese fricatives produced by Japanese speakers including both females and males is less robust than the distinction between the two English fricatives produced by English speakers. They hypothesize that the small vowel inventory of Japanese as well as the phonological constraints on the following vowels may allow listeners to rely more on vocalic cues to distinguish the two fricatives. This hypothesis is supported by a few other studies. Hirai et al. (2005) showed that Japanese listeners tend to rely more on transitional cues to the following vowels. Pinter (2007) also offered explanations of the phonological restrictions based on perceptual factors. The findings of these studies indicate that articulations of Japanese voiceless fricatives require less precision for their places of articulation than those of English fricatives because there are other linguistic cues available for the listeners, which further suggests that there is more room for variability in realization of the fricatives among Japanese speakers.

### 4.2.3. Measurements of Fricatives

Historically, research on fricatives has concentrated on three acoustic attributes of the frication noise: amplitude of the noise, duration of the noise, and spectral properties of the noise. Though

---

22 Holliday (2007) also argues that foreign loanwords, such as ‘sherry [ʃɛri]’ or ‘Shakespeare [ʃɛ:kspija]’, have introduced the contrast before /e/ as well, though examples of [si] are still rare in Japanese today.
extensive research has been done on English fricatives, there are only a few studies investigating Japanese fricatives. In the following sections, previous acoustic studies of fricatives, mainly English fricatives, are reviewed while highlighting different acoustic measurements. Because of the primary focus of the phonetic studies of fricatives, literature reviewed in the first half of the literature review will discuss classification of different groups of fricatives based on the acoustic data. Among those studies, the discussion on /s/ and /ʃ/ is highlighted because of the relevancy to the present study. The latter half of the literature review will focus on sociophonetic studies of fricatives and female gender.

4.2.3.1. Amplitude of Noise

Research investigating amplitude of noise has aimed to classify fricatives. Some studies have measured overall noise amplitude, while others have used relative or normalized amplitude. Studies (Behrens and Blumstein 1988, Fox and Nissen 2005, Jongman et al. 2000, Strevens 1960) using overall noise amplitude have commonly discovered that sibilant fricatives like /s/ and /ʃ/ have a significantly higher amplitude than non-sibilant fricatives like /f/ and /θ/. While no differences were found within the group in some studies (Behrens and Blumstein 1988, Fox and Nissen 2005, Strevens 1960), Jongman et al. (2000) found that /ʃ/ has a greater noise amplitude than /s/. However, in the same study, Jongman et al. report that relative amplitude exhibited a reversal pattern, with /s/ showing higher relative amplitude than /ʃ/. According to Jongman et al., relative amplitude was calculated as the difference between fricative and the following vowel amplitude. Another study (Hedrick and Ohde 1993) that investigated relative amplitude also suggested that /s/ and /ʃ/ can be distinguished by relative amplitude with much greater relative amplitude for /s/ than /ʃ/. 
4.2.3.2. Noise Duration

As is reported by Behrens and Blumstein (1988), noise duration classifies sibilant from non-sibilant fricatives with /s/ and /ʃ/ being longer than /f/ and /θ/, but they did not find any differences in duration between /s/ and /ʃ/. These results are further confirmed by Fox and Nissen (2005), and Gordon, Barthmaier, and Sands (2002; see results for Chickasaw). Jongman et al. (2000) more carefully investigated noise duration using “normalized duration,” defined as the ratio of fricative duration over word duration rather than absolute duration and found a significant difference between /s/ and /ʃ/ in the durations, with /ʃ/ being longer than /s/.

4.2.3.3. Spectral Analysis

With a number of dimensions, a spectrum of the frication noise is the most commonly used means to characterize fricatives. This method of acoustic analysis uses a Fast Fourier Transformed (FFT) spectrum, as is illustrated in Figure 4.2. Many studies have sought to differentiate fricatives across places of articulation by using spectral analysis. One of the most widely used parameters of spectral measurements is spectral peaks which are frequencies with the highest noise amplitudes. It is consistently reported that /s/ has higher spectral peaks than /ʃ/ (Behrens and Blumstein 1988, Seitz and Bladon 1986, Fox and Nissen 2005, Hughes and Halles 1956, Jongman et al. 2000). In a study of the acoustic characteristics of English voiceless fricatives, Behrens and Blumstein (1988) found that /s/ and /ʃ/ can be reliably distinguished by major frequency peaks between 3.5-5k Hz for /s/ and 2.5-3.5k Hz for /ʃ/. Hughes and Halle (1956) also report that the frequency peaks for /s/ are generally situated above 4k Hz, while the spectra of /ʃ/ shows peaks at lower frequencies. While it has been documented that the noise
spectrum of fricatives is influenced by the following vowels (Fujisaki and Kunisaki 1976, Heinz and Stevens 1961, Mann and Repp 1980), Bladon and Seitz (1986) demonstrated that the orientation of the low frequency spectral edge of the first peak can differentiate between /s/ and /ʃ/ regardless of the following vowel contexts.

![FFT spectrum of [s] in “sonna (‘such’)” produced by the Tokyo speaker Hana](image)

Figure 4.2. FFT spectrum of [s] in “sonna (‘such’)” produced by the Tokyo speaker Hana

Another widely used dimension of a spectrum of the fricatives is spectral moments. In the moment analysis approach, the aperiodic energy of fricatives is treated as a random probability distribution, with the first four moments of center of gravity (often called ‘mean’), standard deviation (or ‘variance’), skewness, and kurtosis. With this technique, Forest et al. (1988) aimed to establish a quantitative metric independent of speaker and vowel context. Center of gravity (COG) is the frequency around which there is the greatest concentration (or ‘centroid’) of energy. Standard deviation (SD) measures the degree of frequencies in the spectrum diverging from the COG. Skewness examines the distribution of energy between the higher and lower frequencies, with more negative skewness showing that acoustic energy is concentrated in the higher

---

23 Mann and Repp (1980: 213) explain that the acoustic context dependency is mainly a consequence of anticipatory lip rounding for vowels such as [u] and [o].
frequencies and more positive skewness indicating that acoustic energy is concentrated in the lower frequencies. Kurtosis is a measure of the degree of steepness of the spectrum. A spectrum with flatter, less defined spectral peaks displays a lower kurtosis than a spectrum with high, well-defined peaks (Forrest et al. 1988, Jongman et al. 2000).

Studies using the spectral moments for classification purposes have generally supported that /s/ and /ʃ/ can be differentiated by spectral moments, though the findings are more consistent for COG and skewness. Primarily, /s/ is reported to display higher COG than /ʃ/ (Jongman et al. 2000, Fox and Nissen 2005, Shadle and Mair 1996), and /s/ displays more negative skewness than /ʃ/ (McFarland et al 1996, Jongman et al. 2000). Fewer studies have reported data of SD and kurtosis. Jongman et al. (2000) note that SD is lower in /s/ than /ʃ/. The results of kurtosis are not consistent. Jongman et al. (2000) found that the kurtosis is higher for /s/ than that for /ʃ/, but Nissen and Fox (2005) report a reverse pattern. Koening (2013) points out that these discrepancies may result from differences in analytical procedures such as sampling rate and whether or not spectral averaging was used.

4.3. Background II: Fricatives and Gender

4.3.1. Male-Female Differences

Gender differences in realization of fricatives are reported in previous literature. Women are widely reported to produce /s/ and /ʃ/ at a higher COG and at higher peaks (Jongman et al. 2000, Maniya et al. 2008, Nittroeur et al. 1989 Schwartz 1968, Yeni-Komshian and Soli 1981, Flipsen et al. 1999, Stuart-Smith, Timmins, and Tweddele 2007). Some studies have also reported the tendency of women to have higher SD, more negative skewness, and greater kurtosis as well (Flipsen et al. 1999, Fox and Nissen 2005, Fuchs and Toda 2010, Heffernan
2004, Jongman et al. 2000). Gender differences in fricative duration were also found by Jongman et al. (2000) with /s/ and /ʃ/ in female speech being shorter than those in male speech, though Fox and Nissen (2005) failed to reach significance. These male-female differences are often explained by the physiological differences such as different sizes of the vocal tract between women and men. Because men tend to have larger vocal tract, the acoustic resonances of men’s voices tend to be lower in frequency than those of women’s voices (Avery and Liss 1996, Johnson 1991).

4.3.2. Non-Physiological Patterns

Fuchs and Toda (2010: 284) challenge the biological underpinning of gender differences. By analyzing both acoustic and physiological measures of German and British English speakers, Fuchs and Toda examined whether the gender differences in the realization of /s/ are best accounted for by anatomy. In conclusion, they found support for both anatomical and sociolinguistic explanations. Although their initial speculation was that physiological differences between males and females are larger in the pharynx, (the back part of the vocal tract) than at the front cavity where alveolar fricatives are realized, the size of speakers’ front cavity was found to be a predictor for gender differences among British English speakers. Despite the unexpected correlation, a statistical analysis factoring out potential anatomical influences between British female and male speakers still presented females with higher COG than males. Furthermore, while the overall constriction width of German females was significantly wider than that of British females, the German females realized anterior articulation in a similar manner to the English females.

The anatomical explanation further loses validity when the focus is shifted into sociolinguistic variation within gender categories. For instance, in her work on Glaswegian
English, Stuart-Smith (2007) found that even though there may be physiological differences between younger (13-14 years old) and older (40-60 years old) males, the measured acoustic parameters are relatively similar. By contrast, the young working class girls behave more similarly to the males than other female groups including older working class women, middle class women, and middle class girls. This study sheds light on the fact that construction of gender is more multifaceted than just a simple male-female distinction; constructing orthogonal dimensions of identity such as class and age may contribute as much as using ‘masculine’ or ‘feminine’ features does.

4.3.3. Age

Effects of age, in addition to gender, in realization of fricatives are also examined to explore the age at which the gender differences become evident. Although in his study of children ranging in ages from 3 to 8, Munson (2004) did not find any significant differences between gender in all of the four measures he examined (spectral variability, mean duration, variability in duration, and coarticulation), gender distinction is demonstrated in the four moments of /s/ among 9- to 15-year-olds by Flipsen et al. (1999). Holliday et al. (2010) reported evidence of gender differences for /s/ among girls as young as 3 years old. Using measures of peakERB (spectral peak calculated on a perceptional scale) and category compactness, Holiday et al. derived a measure of the “robustness” of the /s/-/ʃ/ contrast for adults and children aged 2 to 5 years old and represented the accuracy rate with which a fitted logistic regression model predicts the target fricative. The two key findings are that the English-speaking adult female participants show more robust differentiation between /s/ and /ʃ/ than their male counterpart, and that a large difference in robustness in female speech is observed between boys and girls from age 3. The authors interpret this difference as a sign of acquisition of gender-marking.
investigation of Standard Southern British English-speaking adults and children ranging in ages from 9 to 14, Romeo, Hazen, Pettinato (2013) also found that females show greater contrast between /s/ and /ʃ/ in their COG. However, in their study, the gender difference does not emerge until approximately age 11, and 9- and 10-years-olds exhibit no male-female difference in category distance. Considering the gap with the finding of Holliday et al. (2010), Romeo et al. (2013: 3790) speculate that the gender distinctions may become accentuated later in adolescence.

4.3.4. Trans Identities

The feminine indexical meaning of high frequency energy as well as the masculine indexicality of low frequency are further evident in sociophonetic studies of realizations of fricatives among transgender speakers. Hazenberg (2012) analyzed a total of 427 tokens of /s/ in the discourse marker so collected from sociolinguistic interviews with 24 participants (4 “queer” males, 4 “queer” females, 4 trans men (female-to-male transsexuals), 4 trans women (male-to-female transsexuals), 4 straight men, and 4 straight women). The results show that the straight women have the highest mean of COG (7500 Hz) followed by queer women (7230 Hz) and trans women (6610 Hz), while the straight men have the lowest (4940 Hz), and queer men (6290 Hz) and trans men (6135 Hz) fall between cissexual25 women and straight men. The fact that queer women and trans women pattern with straight women rather than straight men indicates that

---

24 Hazenberg (2007: 4) notes that “[q]ueer is the preferred umbrella term for the loose association of non-heteronormative identities in Ottawa” which replaces “the cumbersome and almost immediately out-of-date acronym LGBTQQIA2...” While the acronym “LGBTQQIA2...” includes “[l]esbian, gay, bisexual, transgender, transsexual, queer, questioning, intersex, asexual, two-spirit, etc.,” in his study, Hazenberg distinguishes transsexuals from other “queer” labels by calling them trans and the others queer.

25 Cissexual refers to “someone whose body and internal sense of gender identity are congruent” and includes both “queer” and “non-queer” participants (Hazenberg 2007: 5). In other words, straight women and queer women are considered to be cissexual, as are straight men and queer men. Trans men and women do not fall into the category of cissexual.
/s/ is actively produced with higher frequency energy in feminine gender performance by the “queer” women and trans women. At the same time, a cluster analysis conducted with the same data reveals that the straight men, who exhibit the lowest COG, are removed from all the other groups, which suggests that straight men use low frequency energy to index masculinity while distancing themselves from queer and trans men.

In Zimman’s (2012) longitudinal study, 10 trans men recorded for a year long period exhibited an overall negative relationship between COG for /s/ and number of weeks on testosterone. What is strikingly unique about the trans men in Zimman’s data is that the entire population exhibited a much wider range of variation (4,000 Hz – 11,000 Hz) than both female (6,500 – 8,100 Hz) and male (4,000 – 7,000 Hz) ranges suggested by Flipsen et al. (1999). This finding shows that transmasculine speakers construct their identities “in ways that capture their multifaceted relationships with masculinity” by utilizing various resources available to them whether driven by hormonal changes, childhood language socialization, or gendered habits learned later in life (Zimman 2012: 210).

4.3.5. Cross-linguistic Cases

While the tendency for women to exhibit higher frequency energy than men is highly established, mainly in English, cross-linguistic investigations highlight that variation can be found within gender categories. In Fuchs and Toda (2010), gender differences were found among British English speakers but not among German speakers. Fricatives in seven different languages (Aleut, Apache, Chicksaw, Scottish Gaelic, Hupa, Montana Salish, and Toda) are studied in Gordon, Barthmaier, and Sands (2002), and gender was found to be a significant predictor only in Chicksaw. In Holliday et al. (2010), the distinction between /s/ and /ʃ/ was reported to be more robust in English-speaking adult women than in English-speaking adult men, while Japanese
women in their study do not exhibit a more robust differentiation between /s/ and /ɕ/ than Japanese men. Additionally, in the same study, while the gender distinction is observed among English-speaking children at age 3, there is no difference between male and female 5- and 4-year-old children among Japanese speakers. Heffernan (2004) compared mean values of the four spectral moments of /s/ in the speech of Japanese and Canadian speakers and argued that gender differences are greater in Canadian English speakers’ COG than in that of Japanese speakers. Heffernan speculated that the realization of /s/ is less salient in marking speakers’ gender categories in Japanese than in English is because Japanese has a number of other ways to mark gender including morphological and lexical features.

4.3.6. Japanese Fricatives and Gender

Though Holliday et al. (2010) and Heffernan (2004) concluded that the realizations of fricatives are less robust as a gender marker in Japanese than in English, F. Inoue (1989) discusses the “new trend” of sibilant fronting among young women in the Tokyo area. The read-aloud experiment by F. Inoue and Ogino (1984; cited in F. Inoue 1989) shows that the fronting is more prevalent among women than men and among younger women than among older women. Particularly, adolescent females most actively use fronting. Additionally, in another read-aloud experiment with college students, F. Inoue and Ogino (1985; cited in F. Inoue 1989) found regional variation in fronting among female speakers. The feature was found to be dominantly used in the Tokyo area, whereas it was used the second least frequently in the Kinki region, composed of Kyoto and Osaka, and it was used least frequently in the northern area (the Hokkaido and Tohoku region).26 In F. Inoue (1989), this sibilant is labeled as a “burikko pronunciation.” As is noted in Chapter 1, burikko is a deprecatory term, coined in 1980 by mass

---

26 They also found that the fronting is commonly observed in the Kyushu region (the most southwest island of mainland Japan).
media, that refers to a girl who presents “feigned naïveté” and ultra-femininity (Miller 2004). F. Inoue (1989) used this label (i.e. “burikko pronunciation”) since public discourse in the 1980s often associated sibilant fronting with burikko speech and the association derives from the “lispy” and baby-like sounds of the feature. However, as Miller (2004: 149) points out, F. Inoue (1989) does not offer further evidence for using this particular label other than the fact that the feature is exclusively used by young women.

In addition to the fronting of the place of articulation, Japanese cultural references often discuss the degree of frication. For example, a Japanese linguist Umegaki (1962: 23) impressionistically notes that Western Japanese fricatives are generally “less fricable” than Eastern fricatives. A voice trainer for professional singers and newscasters, Ohshima (2012 November 9), advises her students to “reduce frication” to attain “elegant speech” since strong fricatives are “unpleasant to listen to (mimizawarina)” and “not perceived as “a beautiful sound (utsukushii oto).” Though the relation between the intuitive degrees of frication and their acoustic properties is not very clear, it is significant that the frication noise is somewhat associated with the cultural ideology of locality and possibly gender, since “beauty” of speech is generally considered as a more desirable trait for women than men.

While previous (Western) literature of sibilant production and gender has provided strong evidence for the ideology that associates high frequency energy with femininity, there has been a tendency among many studies to treat femininity as a homogeneous quality. There is virtually no study investigating different types of femininities constructed by speakers through the realization of sibilants. This is because the proposed meanings of varied realizations of sibilants/fricatives do not offer any better explanation beyond physiologically based roots. In this study, I will explore more intuitive indexical meanings of realizations of fricatives. To link the acoustic metrics of fricatives, I will borrow psychoacoustic notions of sound images.
4.4. Psychoacoustic Parameters of Sound Image

Research in psychoacoustics is dedicated to understanding the relationship between acoustic sound signals, the physiology of the auditory system, and the psychological perception of sound. The three basic parameters of sound quality highlighted in this section are *loudness*, *sharpness*, and *clarity*. While these qualities are relevant to sound symbolism of language through *iconic* indexicality in Peirce’s sense, there are only a few linguistic studies that make use of these parameters. These psychoacoustic notions are widely applied outside of linguistics in fields such as engineering to analyze the noise produced by consumer products such as home appliances and automobiles.

*Loudness*

Loudness indicates a subjective perception of the intensity of a sound, and therefore, the acoustic property of *intensity* is relevant to the parameter. The most widely accepted measurement of loudness was originally proposed by Zwicker (1960, 1990), which measures the perceptual loudness level \( N \) (in sones) on the basis of the excitation pattern of the basilar membrane in the ear. The computation of \( N \) uses a fairly complex graphical method based on 1/3-octave band sound pressure levels. The complexity arises because loudness perception is a function of sound pressure level, frequency, and the spectral shape of the sound. Significantly, this model shows that overall intensity is not the only acoustic correlate of loudness; perceived loudness can be calculated by integrating intensity within specific frequency bands (“critical bands”). Whereas the energies in the low frequency bands contribute little to perceived loudness, the energies in the high frequency bands contribute a much greater addition to perceived loudness.
Glave and Rietveld (1975) examined the amount of effort a speaker puts into producing speech and the resulting speech loudness in order to make a comparison between experimental loudness data and calculated loudness data using Zwicker’s model. First, for the “non-effort stimuli,” five speakers were instructed to articulate sustained vowels “as normally as possible” while keeping the pitch at 200 Hz (sinusoidal signal was fed through headphones). For the “effort stimuli,” the speakers were asked to produce vowels “for almost three seconds extremely loudly.” The two recorded stimuli were both adjusted at 65 dB rms and converted into a perceptual unit, and all vowel-dependent and speaker-dependent intensity differences were eliminated. 25 listeners were asked to listen to the stimuli and match the loudness of sinusoidal signal to be “equally loud” as the stimuli by adjusting the volume. Then, the listeners’ perceived loudness (the volume set by the listeners as equally loud as the stimuli) was measured by an intensity meter. Finally, the difference in the loudness between the effort and non-effort stimuli based on the listeners’ loudness-matching procedure was compared to the differences in loudness calculated using Zwicker’s model. They reported that a constant difference existed in the listeners’ perceived loudness and also in the calculated loudness. Furthermore, they calculated the correlation coefficient of the two loudness differences and found that the perceived and calculated loudness correlated to a large extent.

**Sharpness**

Sharpness is a measure of the high frequency content of a sound (Zwicker and Fastl 1990). Spectral peaks, COG, and skewness are relevant spectral metrics to this parameter since spectral peaks are the frequencies with the highest noise amplitudes, and COG and skewness present the frequency levels where the energy is concentrated. In general, sharpness is increased by adding higher frequency content, and decreased by adding lower frequency content (Fastl
The correlation between frequency and sound sharpness is not in direct proportion due to the psychophysical effects of human hearing. The unit of sharpness is the acum, and one acum is defined as the sharpness of a narrow band of noise one critical bandwidth wide at a center frequency of 1 kHz having a level of 60 dB(SPL) (Thorne 2007: 77). Sharpness is closely related to loudness as is clear from the fact that its calculations are based on specific loudness computations. The model underlines that perceived loudness is calculated by integrating intensity within specific frequency bands ("critical bands"). Given the same level of intensity, sharper sounds are perceived louder than less sharp sounds.

The sensation of sharpness is inverse to sensory pleasantness. Fastl (2005: 147) argues that "the right amount of sharpness" can give sound a character of powerfulness, but "too much sharpness" will render a sound aggressive. In this regard, sharpness is a psychoacoustic quality that is often used for examination of the perceived annoyingness of sounds (e.g. Genuit 2001) In a language-related study, Sapp, Wolters, and Becker (1998) discuss sibilant reduction for obtaining speech recording that produces a highly realistic image of the original acoustic event. Due to signal distortions during the recording process, some artifacts may be introduced. One frequently occurring problem is sibilants. To detect the threshold of sharpness that causes disturbance, they asked subjects to rate 141 sibilants occurring in the test sentences from 0.5 (hardly disturbing) to 4 (very disturbing). Based on the sharpness values of those sibilants calculated by Zwicker’s formula, sharpness exceeding 1.2 acum was determined to be perceived as disturbing.

The parameter of sharpness is used in a few linguistic studies of fricatives. Goodacre and Nakajima (2005) used this critical-band scale to examine the differentiability of the English fricatives /s/ and /ʃ/. One subject was asked to provide a perceptual impression for each narrow band of noise that “mimics” the spectral peaks of fricatives. The stimuli were based on Zwicker’s critical-band rate scale, which divides the audible frequency range up to 15,500 Hz into 24 abutting critical bands. Under this scale, for example, the first band ranges from 0 to 100 Hz,
the next from 100 to 200 Hz, and the last 24th band ranges from 12,000 to 15,500 Hz. For the rating, the listener was to allocate points out of a total of 10 to the five English voiceless fricatives (/h/, /f/, /ʃ/, /s/ and /θ/) based on her or his judgment of which fricatives the stimulus evoked. For example, if a stimulus elicited a strong /ʃ/ sensation together with a slight sensation of /s/, then the subject recorded a response of 8 and 2 for /ʃ/ and /s/. The results presented a clear distinction between /s/ and /ʃ/ with /ʃ/ receiving the highest scores between 12th and 18th band ranges and /s/ receiving the highest scores between 19th and 22nd band ranges. By using the psychoacoustic parameter, Goodacre and Nakajima were able to connect the speech perception mechanism to the well-studied contrast between the two English fricatives.

Sharf (1971) also investigated how well subjects categorize English voiceless and voiced consonants on the basis of various perceptual parameters including sharpness derived from the time, intensity, and frequency dimensions of speech. Particularly, the question was how successfully perceptual parameters can predict categorization of those consonants on the basis of manner and place of articulation as well as voicing. Consonant-vowel syllables (/p, t, k, b, p, g, f, s, j, v, z, ʒ/ and /i, a, u/) were first recorded for the stimuli. 16 subjects were instructed to rank each consonant on a 7-point scale according to a different perceptual parameter, including duration (long-short), loudness (loud-soft), frequency (high-low), sharpness (sharp-dull) and contact (front-back). Sharf used the term sharpness in addition to frequency because it is a more commonly used term that refers to the spectral differences among consonants. The overall results indicated that duration is an appropriate perceptual parameter to distinguish consonants with different manners of articulation, while contact successfully differentiates the places of articulation of consonants. Frequency tended to differentiate voiced and voiceless sounds, and sharpness tended to distinguish fricatives from stops. For the /s/-/ʃ/ contrast, the only parameters that showed any significant differences were sharpness and contact. /s/ is perceived as significantly sharper than /ʃ/.
Finally, sound clarity is a term used in sound quality studies that refers to “the perceived resolution of the auditory image, that is the precision with which the details of sound can be heard” (Miskiewicz et al. 2012: 41). The spectral measurements relevant to sound clarity are SD and kurtosis since SD shows dispersions of energy, and kurtosis shows concentration of energy. While the parameter of clarity is often used in the arena of room acoustics such as concert hall designs, when it is applied to speech, sound clarity is a measure of speech intelligibility that indicates the ease of understanding speech. SD and kurtosis are two spectral metrics that are related to this parameter. It is a complex function of psychoacoustics, the signal to noise ration (SNR) of the sound source, and direct-to-reverberant energy within the listening environment (Perez and Rodriguez-Esteban 2004 January 20).

The notion of clarity has been accepted outside of psychoacoustics, and linguists have studied clarity of speech. Acoustic descriptions of clear speech are generally concerned with global (utterance-level) patterns such as slower speech rate, more frequent and longer pauses, higher F0, and wider range of F0 in linguistics (Krause and Braida 2004, Bradlow et al. 2003, Liu et al. 2004, and Smiljanic and Bradlow 2005). However, an exceptional study conducted by Maniwa (2007) investigated speech clarity based on realization of eight English fricatives (/θ, ð, f, v, s, ʃ, z, ʒ/). Maniwa examined realization of the eight fricatives followed by the vowel /a/ in a clear (and careful) speech style recorded in a lab and a non-clear speech style recorded in relaxed conversations and compared various measurements including spectral peaks, and spectral moments. In the results it was found that in clear speech /s/ and /ʃ/ decrease in SD and increase in kurtosis while spectral peaks, COG, and skewness show no changes. Noise duration and F0 also presented significant increases in clear speech. In clear speech, the difference in the
frequency energy between the two fricatives becomes larger with an increase of the frequency of /s/ and a decrease of the frequency of /ʃ/. Furthermore, the speaker ratings obtained in the same study (Maniwa 2007, Maniwa et al. 2008) confirmed the common assumption that clear speech enhances intelligibility.

In this section I have outlined the three sound quality parameters, namely, loudness, sharpness, and clarity and reviewed some linguistic studies that make use of those parameters. These studies show that the psychoacoustic parameters accurately capture listeners' perceptions. In particular, it seems that fricatives can be described in terms of degree of sharpness. Apparently, this parameter offers a more perceptual and intuitive quality of fricatives than the one-to-one link between height of frequency energy and characterizations/categorizations such as female gender or femininity. Now that we have established the theoretical and methodological basis for the studies of fricatives, I turn my attention to the current study.

4.5. Current Study

4.5.1. Introduction

The analyses of the two fricatives in this study exhibit clear contrasts between Tokyo and Osaka, which suggests that the fricatives are associated with the regionality of Osaka and Tokyo depending on the realizations. On the other hand, the realization of the fricatives by the Kyoto speakers do not robustly vary from that of the Tokyo speakers or the Osaka speakers, which suggests that the fricatives do not index regionality per se. I suggest that the variants also carry different meanings in addition to regionality. Some meanings are proposed by drawing on the psychoacoustic literature discussed above.
4.5.2. Procedure

For the analysis of fricatives in this study, the spectral peak and the four spectral moments, which incorporate both local (spectral peak) and more global (spectral shape) information, are measured at the mid point of each segment. The metrics are summarized below. The physical unit of the frequency of spectral peaks and COG was converted into ERB (Equivalent Rectangular Bandwidth), a psychophysical measure proposed by Glasberg and Moore (1990).

**Spectral Peak:** Frequencies with the highest noise amplitudes, which are relevant to sound sharpness

**Center of Gravity (COG):** The frequency around which there is the greatest concentration of energy, which is relevant to sound sharpness

**Standard Deviation (SD):** The range of frequencies in the distribution diverging from the COG which is relevant to sound clarity

**Skewness:** The distribution of energy (larger and positive values indicates energy concentration in the lower frequencies) which is relevant to sound sharpness

**Kurtosis:** Peakedness of the distribution (high values show well-defined peaks, indicating more energy is concentrated around the point) which is relevant to sound clarity

Additionally, F2 of the preceding and following sounds were measured for each segment to examine the effect of co-articulation. Intensity of the fricatives was also measured and normalized as *relative intensity*. Relative intensity was calculated as the difference between fricative intensity and following vowel intensity.

Initially, 100 tokens were coded for both /s/ and /ɕ/ for each speaker, but the coding was organized for the following reasons. Tokens where the preceding or following sounds are devoiced as well as where the boundaries with the preceding or following sounds are unclear for any reasons (e.g. unclear articulation, as determined by my auditory analysis while confirming with the spectrograms) are excluded from the data. IP (intonational phrase) initial tokens were
eliminated since the IP initial /s/ tends to exhibit higher frequency energy than non-IP initial /s/ (Podesva 2013). Tokens that are adjacent to geminates were excluded since gemination affects F1 of the following sounds (Kawahara 2006). Any tokens that are vocalized were also eliminated. Finally, tokens with measurement failures where no values were recorded by PRAAT (Boersma and Weenink 2013) were deleted from the coding. After the cleanup of the data, the total number of tokens for /s/ was 3,525 (Tokyo N=1,194, Kyoto N=1,175, Osaka N=1,156; average N/participant =76) and for /ɕ/ was 3,376 (Tokyo N=1,030, Kyoto N=1,280, Osaka N=1,066; average N/participant = 73).

For measurement, I used a PRAAT script that was prepared by modifying the script used for Stanford University’s Voices of California project. For each interval labeled as /s/ in the text grids the measurements were made at the mid point. Spectral peak location of the fricatives was examined using a 40 ms. full Hamming window placed at the mid point. Then, the script defined and recorded the spectral peak frequency as the highest-amplitude peak of the FFT spectrum. For measurements of the four spectral moments, the script employed the PRAAT automated moments analysis function. The same procedure was repeated for the intervals labeled as /ɕ/. The spectral peak and COG values recorded by the PRAAT script were manually converted into ERB after the measurement process was completed.

Mixed effects liner regressions were conducted using R, with the spectral peaks, the four spectral moments, and relative intensity as the dependent variables. F2 of the preceding and following vowels in addition to region were treated as the independent variables. Speaker was included as a random effect.

4.5.3. Results
Results of the regression analyses are summarized in the tables below. The intercept indicates the estimated values of the spectral measures (i.e. Spectral Peaks, COG, SD, Skewness, Kurtosis, Intensity) if all independent variables (PreF2 (F2 of the preceding sound), FolF2 (F2 of the following sound), and Region) are the defaults. In these particular models, Tokyo is treated as the intercept.

Tables 4.1-4.6 present results of the statistical analysis for /s/, and Figure 4.3 displays the cross-regional variation in the mean values and the degrees of dispersion. The regression models found region to be a significant predictor in realization of /s/ for the four spectral moments but not for spectral peaks and intensity. As Tables 4.1 shows, values of the spectral peaks are not statistically different across the three regions. Comparisons between the Tokyo /s/ and the Osaka /s/ present significant differences in measurements of all of the four spectral moments. The Osaka /s/ shows higher COG (Table 4.2), higher SD (Table 4.3), lower skewness (Table 4.4), and lower kurtosis (Table 4.5) than the Tokyo /s/. In the comparisons between the Tokyo /s/ and the Kyoto /s/, significant differences are found only in the metrics of skewness and kurtosis: Kyoto /s/ exhibits lower skewness (Table 4.4) and lower kurtosis than the Tokyo /s/ (Table 4.5). No statistical difference is found among the regions for the intensity of /s/, though F2 of the following sounds is a significant predictor (Table 4.6).

| Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|----------|----------|------------|------------|-------|---------|
| (Intercept) | 7.6427 | 7.6405 | 5.5205 | 9.8559 | 0.0001 | 0.0000 |
| PreF2 | 0.0000 | 0.0000 | -0.0008 | 0.0008 | 0.9998 | 0.9936 |
| FolF2 | -0.0001 | -0.0001 | -0.0010 | 0.0008 | 0.8458 | 0.8494 |
| Regionkyoto | 1.4332 | 1.4500 | -0.5756 | 3.5283 | 0.1572 | 0.1807 |
| Regionosaka | 0.5003 | 0.4994 | -1.4770 | 2.5827 | 0.6234 | 0.6358 |

Table 4.1. Mixed effects for model of Spectral Peaks (ERB) for /s/
|                  | Estimate  | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|)  |
|------------------|-----------|----------|------------|------------|-------|----------|
| (Intercept)      | 22.6823   | 22.6580  | 20.8024    | 24.5411    | 0.0001| 0.0000   |
| PreF2            | 0.0002    | 0.0003   | -0.0003    | 0.0008     | 0.3900| 0.3962   |
| FolF2            | -0.0004   | -0.0004  | -0.0010    | 0.0002     | 0.2244| 0.2174   |
| Regionkyoto      | 2.1176    | 2.1254   | -0.0602    | 4.2469     | 0.0540| 0.0826   |
| Regionosaka      | 2.6432    | 2.6529   | 0.5474     | 4.7100     | 0.0116| 0.0279*  |

Table 4.2. Mixed effects for model of COG (ERB) for /s/

|                  | Estimate  | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|)  |
|------------------|-----------|----------|------------|------------|-------|----------|
| (Intercept)      | 3303.7945 | 3302.6766| 3031.6510  | 3564.5526  | 0.0001| 0.0000   |
| PreF2            | -0.0028   | -0.0024  | -0.0925    | 0.0802     | 0.9508| 0.9495   |
| FolF2            | -0.0201   | -0.0199  | -0.1160    | 0.0755     | 0.6788| 0.6849   |
| Regionkyoto      | 188.2425  | 189.6193 | -93.3928   | 480.7724   | 0.1878| 0.2332   |
| Regionosaka      | 380.7116  | 379.5430 | 98.6734    | 664.1621   | 0.0106| 0.0145*  |

Table 4.3. Mixed effects for model of SD for /s/

|                  | Estimate  | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|)  |
|------------------|-----------|----------|------------|------------|-------|----------|
| (Intercept)      | 2.0317    | 2.0293   | 1.2036     | 2.8821     | 0.0001| 0.0000   |
| PreF2            | 0.0001    | 0.0001   | -0.0003    | 0.0004     | 0.7380| 0.7146   |
| FolF2            | 0.0001    | 0.0001   | -0.0002    | 0.0005     | 0.5208| 0.5295   |
| Regionkyoto      | -0.8782   | -0.8709  | -1.6383    | -0.0785    | 0.0298| 0.0268*  |
| Regionosaka      | -0.9542   | -0.9513  | -1.6939    | -0.1685    | 0.0184| 0.0148*  |

Table 4.4. Mixed effects for model of Skewness for /s/

|                  | Estimate  | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|)  |
|------------------|-----------|----------|------------|------------|-------|----------|
| (Intercept)      | 13.2109   | 12.9915  | -35.2868   | 59.6043    | 0.5922| 0.5814   |
| PreF2            | 0.0112    | 0.0112   | -0.0112    | 0.0334     | 0.3198| 0.3225   |
| FolF2            | 0.0038    | 0.0041   | -0.0211    | 0.0287     | 0.7504| 0.7663   |
| Regionkyoto      | -27.0115  | -27.2223 | -49.2062   | -5.3232    | 0.0154| 0.0102*  |
| Regionosaka      | -25.5187  | -25.5810 | -47.4401   | -3.3989    | 0.0208| 0.0156*  |

Table 4.5. Mixed effects for model of Kurtosis for /s/
|                | Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|----------------|----------|----------|------------|------------|-------|----------|
| (Intercept)    | -13.1866 | -13.2070 | -14.2233   | -12.1241   | 0.0001| 0.0000   |
| PreF2          | 0.0002   | 0.0002   | -0.0002    | 0.0006     | 0.2308| 0.2422   |
| FolF2          | 0.0007   | 0.0007   | 0.0003     | 0.0011     | 0.0014| 0.0014** |
| Regionkyoto    | -0.1344  | -0.1264  | -1.1170    | 0.8557     | 0.7976| 0.7946   |
| Regionosaka    | 0.4247   | 0.4390   | -0.5386    | 1.4142     | 0.3666| 0.4046   |

Table 4.6. Mixed effects for model of Intensity for /s/

Figure 4.3. Box plots of the measurements for /s/
Tables 4.7-4.11 below present results for /ɕ/, and Figure 4.4 provides box plots of the cross-regional variation. Fewer significant differences are found in /ɕ/, but the measurements that show differences have the same patterns with those of /s/. Region is a significant predictor in the spectral measurements only between Tokyo and Osaka, and not Tokyo and Kyoto. The Osaka /ɕ/ exhibits significantly higher COG (Table 4.7) and lower skewness (Table 4.9) than the Tokyo /ɕ/, which are both identical patterns with those of /s/. Finally, F2 of the both preceding and following sounds serve a significant predictor for the intensity of /ɕ/ (Table 4.11).

|              | Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|--------------|----------|----------|------------|------------|--------|----------|
| (Intercept)  | 11.2312  | 11.2314  | 8.4224     | 14.0477    | 0.0001 | 0.0000   |
| PreF2        | 0.0001   | 0.0001   | -0.0007    | 0.0010     | 0.7462 | 0.7666   |
| FolF2        | 0.0000   | 0.0000   | -0.0008    | 0.0009     | 0.9790 | 0.9952   |
| Regionkyoto  | -0.1825  | -0.1996  | -2.8431    | 2.5553     | 0.8828 | 0.8978   |
| Regionosaka  | -1.1888  | -1.1982  | -3.9315    | 1.3626     | 0.3578 | 0.4068   |

Table 4.7. Mixed effects for model of Spectral Peaks (ERB) for /ɕ/

|              | Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|--------------|----------|----------|------------|------------|--------|----------|
| (Intercept)  | 26.2666  | 26.2582  | 24.7699    | 27.7914    | 0.0001 | 0.0000   |
| PreF2        | 0.0001   | 0.0001   | -0.0003    | 0.0005     | 0.6292 | 0.6282   |
| FolF2        | 0.0001   | 0.0001   | -0.0003    | 0.0005     | 0.7494 | 0.7499   |
| Regionkyoto  | -0.9160  | -0.9204  | -2.4817    | 0.7280     | 0.2582 | 0.3062   |
| Regionosaka  | -2.2738  | -2.2681  | -3.8713    | -0.6255    | 0.0056 | 0.0115*  |

Table 4.8. Mixed effects for model of COG (ERB) for /ɕ/
| Estimate | MCMCmean | HPD95lower | HPD95upper | pMCMC | Pr(>|t|) |
|----------------------|-------------------------------|----------------------|----------------------|----------------------|----------------------|
| (Intercept)          | 2843.1130                     | 2843.6911            | 2639.3784            | 3040.1565            | 0.0001               | 0.0000               |
| PreF2                | -0.0390                       | -0.0385              | -0.0943              | 0.0178               | 0.1622               | 0.1603               |
| FolF2                | 0.0321                        | 0.0320               | 0.0836               | 0.2190               | 0.2211               |
| Regionkyoto          | 57.9193                       | 56.5979              | -156.6406            | 265.9939             | 0.5994               | 0.6226               |
| Regionosaka          | 194.8815                      | 191.8736             | -21.2363             | 401.8063             | 0.0714               | 0.0939               |

Table 4.9. Mixed effects for model of SD for /ɕ/
The general patterns of the seemingly complicated results can be summarized as follows: the Osaka fricatives exhibit 1) higher COG, 2) higher SD, 3) lower skewness, and 4) lower kurtosis than the Tokyo fricatives. The Kyoto fricatives fall between Osaka and Tokyo in COG and SD but behave like Osaka in skewness and kurtosis only in /s/. No cross-regional variation has been found in spectral peaks and intensity.
Other than the factor region, F2 values of the preceding and following sounds were chosen to be a significant predictor for the intensity of the fricatives. As Tables 4.6 and 4.12 show, there is a correlation between F2 and the intensity of /s/ and /ɕ/. The intensity tends to increase when F2 values of the preceding or following vowels increase, which means that the intensity is higher with fronter vowels.

### 4.5.4. Discussion

The overall results of the regression analyses have shown a clear contrast between Tokyo and Osaka. It is clear that the fricatives serve as a regional marker that distinguishes one from the other. Low COG, low SD, high skewness, and high kurtosis characterize the Tokyo regionality, while high COG, high SD, low skewness, and low kurtosis characterize the Osaka regionality. However, being in between, the Kyoto regionality is represented less clearly by these spectral measurements of the fricatives. This suggests that realization of the fricatives does not mark
regionality per se. If the variants are not all about region, the question would be what the variation means. To understand the social meanings of the various realizations of the fricatives, the psychoacoustic notions come into play.

When interpreted in terms of psychoacoustic sound quality, spectral peaks, COG, and skewness are relevant to sound *sharpness*, SD and kurtosis are relevant to sound *clarity*, and intensity is relevant to sound *loudness*. Intensity of the fricatives does not vary cross-regionally (see Tables 4.6 and 4.12), which means that there is no significant difference in sound loudness across the regions. On the other hand, the significantly higher COG and the lower skewness of the Osaka fricatives indicate that the energy is concentrated at higher frequencies in Osaka than it is in Tokyo. Though the correlation between height of frequency energy and sound sharpness is not in direct proportion due to the psychophysical effects of human hearing, these results indicate that the fricatives produced by the Osaka participants are sharper than those produced by the Tokyo participants. Additionally, though intensity has not exhibited significant differences across the regions, sound sharpness is closely related to sound loudness. Given the same intensity level, sounds with higher frequency energy are perceived as louder than sound with lower frequency energy, which suggests that the fricatives of Osaka are also perceived to be louder than the fricatives of Tokyo.

Recall that this parameter of sharpness is often discussed in terms of sensory pleasantness. According to Fastl (2005: 147), “the right amount of sharpness” can imbue sound with a character of powerfulness, but “too much sharpness” will render a sound aggressive. For this reason, fricatives are sometimes considered to be “annoying” and “disturbing” (e.g. Sapp, Wolters, and Becker 1998, Ohshima 2012 November 9). Though measuring the exact degrees of perceptual sharpness using different calculation methods or through perception experiments is not within the scope of the present study, it is suggested here that the perceived sharpness of the fricatives of Osaka corresponds with the harsh, tough, and vulgar image of the Osaka variety.
Meanwhile, the other group of spectral moments is composed of SD and kurtosis. SD shows dispersion of energy, and kurtosis shows concentration of energy. The lower SD and the higher kurtosis of Tokyo shows that the energy of the Tokyo fricatives is less dispersed and more concentrated at one point compared to that of the Osaka fricatives. Less dispersion and more concentration of energy are attained by more precise articulation by the Tokyo speakers. I also argue that clarity of speech corresponds with the “standard” image of the Tokyo variety. Although the association between clarity and standardness has not been previously noted in studies of Japanese, Ladegaard’s (2001) perception study of Standard Danish demonstrates that a clear and precise articulation is a significant factor for listeners to define Standard Danish.

4.6. Conclusion

This chapter has investigated variation in realization of fricatives by female speakers from Tokyo, Kyoto, and Osaka through analysis of the spectral measurements (spectral peaks, COG, SD, skewness, and kurtosis), and intensity. The results have shown that Osaka exhibits higher COG, higher SD, lower skewness, and lower kurtosis than Tokyo, while Kyoto exhibits lower intensity than both Tokyo and Osaka. To understand these acoustic measurements more intuitively, I have made use of the psychoacoustic notion of sound loudness, sharpness, and clarity. While no significant difference in loudness has been found across the regions, the Osaka fricatives index sharpness with higher frequency energy shown with higher COG and lower skewness. Since sharpness and loudness are closely related, though the correlation is not in direct proportion, this finding also suggests that the fricatives produced by the Osaka participants are also perceived to be louder than the fricatives produced by the Tokyo participants. On the other hand, with the lower SD and higher kurtosis, the fricatives produced by the Tokyo speakers connote clarity.
Finally, these characteristics of the fricatives in Tokyo and Osaka correspond with the stereotypical images of the two varieties. The sharpness of the Osaka fricatives, along with the perceived loudness, synchronizes with the “harsh” and “vulgar” Osaka image. The clarity of the Tokyo fricatives is aligned with the “standard” Tokyo image. The sharpness of the high frequency energy of /s/ and its usage by the Kyoto speakers will be further investigated within ethnographic contexts in Chapter 6.
CHAPTER 5
Deaccented Intonation: The Spice of Locality

5.1. Introduction

This chapter investigates intonational patterns of Tokyo, Kyoto, and Osaka, particularly in the negative form of polar (or ‘yes/no’) questions (NPQs (negative polar questions)). Intonational patterns of Japanese NPQs have been discussed in a number of recent studies (e.g. Hara and Kawahara 2012, Inoue 2008, Tanaka 2010, Wakuda 2003) which all highlight the newly emerged type of rising intonation. This phenomenon is named the *tobihane* ‘jumping-up’ tone by Tanaka (1993) because of the shape of the pitch track and more appropriately called “deaccentuation” by Hara and Kawahara (2012) since conventional pitch accents are deleted in the form. Figures 5.1 and 5.2 present examples of an NPQ [*kore] takaku nai?* (‘Isn’t this expensive?’) adapted from Hara and Kawahara (2012). While Figure 5.1 shows a conventional intonational pattern of the NPQ where the lexical accent is retained, Figure 5.2 shows the deaccented intonational pattern.

Figure 5.1. Conventional accented pattern (Adapted from Hara and Kawahara 2012)
In Figures 5.1 and 5.2, the tone labeling is presented in the first tier. “H*+L” stands for the only possible Japanese pitch accent, where the letters (H and L) indicate the high or low pitch and an asterisk indicates the accent. Japanese accent is manifested simply as a sharp fall in pitch. The “%” diacritic shows the boundary tone with %L being the initial boundary tone and L% and H% the final boundary tones. The H tone with a hyphen (“H-”) is called the phrasal high, which serves as the target tone of a gradual rise in pitch starting from %L. Whereas the pitch in Figure 5.1 falls sharply as H*+L after a gradual rise to H-, the pitch in Figure 5.2 continues to rise after it hits the H target tone until the end of the utterance and never presents a sharp fall. While more details about the tone labeling will be discussed later in this chapter, it should be noted here that the latter phenomenon where we observe a continuous rise in pitch without a sharp fall is called “deaccentuation” in this chapter.

This new type of intonational pattern is reported to be used predominantly by young speakers from the Tokyo area (Tanaka 2010, F. Inoue 1998), but the discursive roles of the feature have not been theoretically or sociolinguistically scrutinized. This chapter explores meanings of the feature in gender construction and locality through quantitative and qualitative
analyses, while highlighting the unexpected use of the feature in non-Tokyo regions. Based on the sociolinguistic, pragmatic, and semantic literature on NPQs, rising intonation, and deaccentuation, I demonstrate that in Tokyo, the deaccentuation of NPQs displays a specific interactional stance where the speaker’s epistemic and evidential biases are asserted with a lack of commitment as well as a lack of vulnerability to disagreements, which constructs “urban cuteness.” While cuteness has been often discussed in relation to weakness, a lack of confidence, and intimacy in previous literature (e.g. Kinsella 1995, Miller 2004, Moon 2010, Kajino 2011), the key components of this new type of cuteness are confidence and aloofness. Due to the regional indexicality attached to the form, I further argue that the feature additionally represents another meaning in Kyoto and Osaka, namely the Tokyo-centric ideology. This ideology revolves around “urban cuteness,” which is portrayed in the mass media, and depicts an “ideal” young female figure that is urban, trendy, stylish, style-conscious, and cute just like the “typical” young Tokyo girl. The discussion section highlights that those Kyoto and Osaka users of deaccentuation are devoted consumers of mass media, and close examinations of their interactions illuminate that their use of the deaccented form is symbolic of their alignment with the Tokyo-centric gendered ideology circulated by the mass media.

Before discussing the particular feature, I will first outline the basic background of Japanese intonational systems and dialectal differences. Once a basic understanding of Japanese intonational system is established, I will review some semantic and pragmatic foundations relevant to the deaccented NPQs.

5.2. Background I: Japanese Intonational Systems

5.2.1. Tokyo Japanese Prosodic Organization
Although Japanese has long been classified as a mora-timed language, in opposition to stress-timed languages like English, this classification is controversial.\(^{27}\) One common argument to support the mora-timed system is that moras in Japanese are timing units of roughly constant duration (e.g. Han 1962, Homma 1981), but other phonetic studies denied that any such acoustic temporal unit can be found (e.g. Beckman 1982, Arai 1999, Warner and Arai 2000). Beckman (1982) argued that the mora does not have any phonetic basis, and it is merely a phonological unit of length or weight without phonetic reality. Warner and Arai (2000) agreed that the mora does not show a tendency toward regular timing, but they concluded that it is relevant in the segmentation of Japanese rhythm because of phonological and structural factors.

Japanese prosodic organization has been studied systematically for quite a while and various frameworks have been proposed (e.g. McCawley 1968, Poster 1984, Kubozono 1993, Kawahara and Shinya 2008), with the primary differences among them lying in the levels of prosodic phrasing and terminologies used for those levels. J-ToBI (Japanese Tone and Break Indices) is one of the frameworks that was established to annotate Tokyo Japanese, which I follow to provide the overview of Japanese prosody here. Just like English, Tokyo Japanese is considered to be a pitch accent language, in which the intonational system uses pitch to mark certain moras/syllables in the speech stream. However, Tokyo Japanese and English contrast in the fundamental systems and roles of pitch accents. First, English pitch accent is a post-lexical property, where an interaction of various factors related to the syntax, semantics, pragmatics, discourse, and other sociolinguistic factors determine where the pitch accents are to be placed.

\(^{27}\) Pike (1945) originally proposed that any spoken languages can be categorized as having either stress-timing or syllable-timing. In stress-timed languages like English, German, Russian, and Arabic, stressed syllables in an utterance occur at approximately the same intervals and the time taken for the utterance is proportional to the number of stressed syllables regardless of the number of unstressed syllables, while in syllable-timed languages such as French, Telugu, an Yoruba, each syllable in an utterance bears an approximately equal rhythmic beat, and the amount of time taken for producing the utterance is proportional to the number of syllables. Later, Bloch (1950) argued that languages like Japanese, Ancient Greek, and Sanskrit does not use units of syllables; instead, they use moras, which are traditionally considered as sub-units of syllables consisting of one short vowel and any preceding onset consonants.
In contrast, Japanese pitch accent is a lexical property of a word, which means that the presence or absence of an accent on a particular mora in Japanese speech is predetermined by the given word. Also, while there are a number of pitch accent shapes in English (Pierrehumbert 1980), there is only one type of pitch accent in Japanese, namely H*+L, where the asterisk indicates the pitch accent, and the letters (H or L) represent the high or low pitch. This pitch accent is realized as a sharp fall from a high, occurring near the end of the accented mora, to a low in the following mora.

Venditti (2005) argues that the fact that pitch accent is a lexical property and that there is only one type of pitch accent in Japanese leaves limited domains for variability in distribution of accents based on various paralinguistic meanings and therefore, the Japanese pitch accent itself does not index social meaning. Different meanings associated with different shapes of English pitch accents are discussed at length by Pierrehumbert and Hirschberg (1990). Pierrehumbert and Hirschberg (1990: 285) propose that in English, “speakers use tune to specify a particular relationship between the ‘propositional content’ realized in the intonational phrase over which the tune is employed and the mutual beliefs of participants in the current discourse.” For example, they demonstrate that the H* accents in utterances in general convey that the items made salient are to be treated as “new” in the discourse and that the L*+H accent can convey the impression of uncertainty and lack of speaker commitment to the statement. However, Japanese pitch accent, being available in only one shape that is lexically determined, offers little room for establishing pragmatic meaning attached to the pitch accent.

In Japanese speech, words with particular lexical pitch accents are grouped into prosodic phrases at two different levels. The lower level of prosodic phrasing is called the accentual phrase (AP), which is defined by tone and the degree of perceived disjuncture between words and between groups. This level of a prosodic phrase roughly but not completely corresponds with what are called “minor phrases” (McCawley 1968). APs are typically characterized in Tokyo
Japanese by an initial rise to high tone toward the second mora, followed by a gradual fall to a low tone at the end of the phrase, which means that an AP has at most one pitch accent. In Tokyo Japanese, accented words commonly construct accentual phrases with adjacent (unaccented) words that are usually functional words (such as particles). In less common cases, the adjacent words may also be accented, and in that case, the leftmost accent survives while subsequent accents in the phrase are deleted (Venditti 2005). Accordingly, an AP marks the prosodic grouping itself, which is separate from the contribution of a pitch accent.

The upper level of prosodic grouping is an intonational phrase (IP), which consists of one or more APs. Within an IP, the pitch accent of each AP is retained. Like APs, this level of prosodic phrasing is also defined by the tone and the degree of perceived disjuncture between the groups, but the disjuncture between sequential words across IP boundaries is larger than between words within or across AP boundaries. Also, an IP takes a more significant role in the prosodic domain since it specifies the local pitch range at the beginning of every phrase, which is independent from the ranges set by the former phrases. In Tokyo Japanese, every IP begins with the L pitch and reaches to the H at the pitch accent of the first AP. Then, there is an effect of downstep, by which the pitch height of an AP is compressed when the AP follows another AP within the same IP. This downstep effect is reset at IP boundaries.

---

28 Though this view is found in some studies (e.g. Poster 1984), many other studies including more recent ones (e.g. Maekawa and Igarashi 2007, Vance 2009) disagree with this view since lexically accented particles can form their own APs. See Igarashi (forthcoming) for a more detailed discussion of “deaccenting of particles.”
Figure 5.3. Fo contour and J-TOBI transcription of the utterance “koreto kuroku nai? (‘Isn’t this, for example, dark?’)” by the Tokyo participant Aika.

Figure 5.3 shows an example of the Tokyo participants’ utterances containing an NPQ. The prosodic phrasing of this utterance can be represented as \([\text{AP} \text{Koreto’ka}] \text{[AP Kuroku nai?]}\]. This chapter is concerned with prosodic phrases at the AP level, rather than the IP level, since the target form of the deaccented NPQs occurs as an AP as shown in the example. The deaccented NPQs occur within APs.

Though the sole shape of Japanese lexical pitch accent does not index social meaning as is noted by Venditti (2005) earlier, there are multiple types of boundary pitch movements (BPMs), which usually occur at the right edge of the intonational phrase. Maeda and Venditti (1998) examined a number of rising BPM types in perception and production studies and
concluded that the various BPMs in Tokyo Japanese not only cue differences in pragmatic meaning, but can be also differentiated by F0 height, rise shape, and timing characteristics. In more recent work, Igarashi (forthcoming) lists the following inventory of the boundary tones and summarizes the pragmatic meanings proposed in previous literature:

<table>
<thead>
<tr>
<th>BPM</th>
<th>Description</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>H%</td>
<td>Simple rise</td>
<td>Insisting, persuading, seeking approval, inviting attention, blaming</td>
</tr>
<tr>
<td>LH%</td>
<td>Scooped rise</td>
<td>Expressing intimacy</td>
</tr>
<tr>
<td>HL%</td>
<td>Rise-fall</td>
<td>Explanatory, emphatic</td>
</tr>
<tr>
<td>HLH%</td>
<td>Rise-fall-rise</td>
<td>Infant-directed speech</td>
</tr>
</tbody>
</table>

While the inventory appears to be as simple as only four different types of rising BPMs, Venditti explains that the difference between rises within each category is attributed to differences in various factors like pitch range and voice quality, which cannot be appropriately represented by a J-ToBI annotation system, even though these within-category differences connote different meanings. This point enhances my argument that it is critical to consider co-occurring features in discussion of meaning of variation since linguistic variants carry different, or sometimes competing, meanings when in different combinations of co-occurrence with other features.

Boundary tones are a particularly interesting locus to examine in investigation of questions with conventional intonational patterns since they serve as the only clue to distinguish interrogatives from declaratives in Japanese. Conventionally, there is no other intonational cue to indicate Japanese utterances as interrogatives. However, as is seen clearly from the shape of the pitch track in Figure 5.2 compared to Figure 5.1, a deaccented question displays a contrastive F0 contour to the pitch contour of an accented question. A continuous rise can be observed from the beginning to the end of the utterance. The boundary tone of the deaccented
NPQs is, in its default, H% because deletion of the pitch accents simply prevents the L tone from occurring. The present study focuses on whether the conventional accents are retained or deleted in the APs where the boundary tone simply provides a final pitch target (i.e. H% is the default), and is therefore not concerned with the pragmatic meaning of the particular boundary tone.

5.2.2. Map of Japanese Regional Intonational Patterns

While J-TOBI was established based on the investigation of Tokyo tone phonetics by Pierrehumbert and Beckman (1988), substantial regional variation can be observed in intonational patterns across Japan. Intonation is a feature that survived governmental language standardization efforts and the original forms were maintained more than other lexical, grammatical, and phonetic features. Investigating the status of language standardization, Yoneda (1997) analyzed various phonetic elements and concluded that the standardization of intonational patterns is progressing more gradually than that of the phonemes.
Figure 5.4 shows a Japanese map of regional intonational systems proposed by Kindaichi (2001) based mainly on large-scale surveys of lexical pitch accents. The intonational systems of Japan can be categorized into two major groups: the Tokyo system and the Kyoto-Osaka system. With slight variation within a group, the areas highlighted in mostly white and lettered between A and G all represent the Kyoto-Osaka intonational system. On the other hand, the areas shaded in dark that are numbered between 1 and 6 all represent the Tokyo system. Just like the map of dialectal divisions seen in Chapter 1, Kyoto and Osaka are located in the same intonational region, whereas Tokyo belongs to the other group. While the Tokyo system spreads throughout
the country from the North to the Southwest, the Kyoto-Osaka system is clustered in one relatively smaller area.

5.2.3. Japanese Dialectology and Intonation

Intonation has been widely investigated by Japanese dialectologists. Because Japanese pitch accent is a lexical property, many studies have focused on lexical pitch accents. In most of those traditional studies, researchers ask respondents to pronounce words composed of a different number of moras and make notes of the accentual patterns. Tables 5.1 and 5.2 show the accentual types of the Tokyo system as well as the Kyoto-Osaka system based on Kibe (2010). The number of moras in words groups different accentual patterns. Asterisks indicate the pitch accent, and each letter (H or L) represents one mora with its specified pitch. In both Tokyo and Kyoto-Osaka accentual systems, an accent is realized by a pitch fall. H*+L in Table 5.2 shows that the pitch accent is realized within a mora. Parentheses show that there is no lexical item applicable to the patterns. Note that Tables 5.1 and 5.2 adapt the representation commonly used in traditional Japanese dialectal studies where each mora is labeled with H or L, and the representation is independent from the J-TOBI system.

Tables 5.1 and 5.2 exhibit contrasts between the two accentual systems. As the tables make clear, the Kyoto-Osaka system, having more patterns, is more complicated than the Tokyo system. While the initial pitch of the Tokyo patterns is always low unless the lexical accent is on the initial mora, the Kyoto-Osaka patterns distinguish between high-beginning words and low-beginning words. The contrast between high- or low-beginning words is lexically determined in Kyoto-Osaka Japanese. Another difference in the Kyoto-Osaka system from the Tokyo system is that utterance-final accents are realized as an early fall that is completed within the single mora.
Also, as is indicated by parentheses, high-beginning words never have final accent, and low-beginning words never have initial accent.

<table>
<thead>
<tr>
<th>1 mora</th>
<th>2 moras</th>
<th>3 moras</th>
<th>4 moras</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>L H</td>
<td>L H H</td>
<td>L H H H</td>
</tr>
<tr>
<td>e.g. Ha ‘leaf’</td>
<td>Hana ‘nose’</td>
<td>Sakura ‘cherry tree’</td>
<td>Tomodachi ‘friend’</td>
</tr>
<tr>
<td>H*</td>
<td>L H*</td>
<td>L H H*</td>
<td>L H H*</td>
</tr>
<tr>
<td>Ha ‘tooth’</td>
<td>Hana ‘flower’</td>
<td>L H* L</td>
<td>L H* L</td>
</tr>
<tr>
<td></td>
<td>Otoko ‘male’</td>
<td>Hana ‘flower’</td>
<td>H* L L</td>
</tr>
<tr>
<td></td>
<td>Sakura ‘cherry tree’</td>
<td>H H H H</td>
<td>L H* L</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Imooto ‘younger sister’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mizuumi ‘lake’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Irogami ‘colored paper’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kabuto ‘helmet’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kamakiri ‘mantis’</td>
</tr>
</tbody>
</table>

Table 5.1. Tokyo accentual patterns

<table>
<thead>
<tr>
<th>Initial pitch</th>
<th>1 mora</th>
<th>2 moras</th>
<th>3 moras</th>
<th>4 moras</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>H</td>
<td>H H</td>
<td>H H H</td>
<td>H H H H</td>
</tr>
<tr>
<td>Ha ‘leaf’</td>
<td></td>
<td>Hana ‘nose’</td>
<td>Sakura ‘cherry tree’</td>
<td>Tomodachi ‘friend’</td>
</tr>
<tr>
<td></td>
<td>H*+L</td>
<td>(H H*+L)</td>
<td>(H H H*+L)</td>
<td>(H H H H*+L)</td>
</tr>
<tr>
<td>Hi ‘day’</td>
<td>H* L</td>
<td>H H* L</td>
<td>H H H* L</td>
<td>H H H* L</td>
</tr>
<tr>
<td></td>
<td>Hana ‘flower’</td>
<td>Futari ‘2 people’</td>
<td>Kaminari ‘thunder’</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H* L L</td>
<td>Otoko ‘male’</td>
<td>H H* L L</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Usaki ‘rabbit’</td>
<td>H* L L</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L L H*+L</td>
<td>H H* L</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L H* L</td>
<td>Ninjin ‘carrot’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L H*+L</td>
<td>(L L H*+L)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Macchi ‘match’</td>
<td>L L H* L</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L H* L</td>
<td>Ninjin ‘carrot’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L H* L</td>
<td>(L L H*+L)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L H* L</td>
<td>Ninjin ‘carrot’</td>
</tr>
</tbody>
</table>

Table 5.2. Kyoto-Osaka accentual patterns

156
Although some accentual patterns are commonly observed both in Tokyo and Kyoto-Osaka, their lexical contrasts can be different in the two systems. Fujisaki et al. (1976) summarize examples of accentual patterns of bi-moraic words in Tokyo and Kyoto-Osaka. As is seen in Tables 5.1 and 5.2, the L H*+L and H H types do not occur in Tokyo while all four types occur in Kyoto-Osaka (presented as the Osaka system in the table). Also, the accentual contrast of the minimal pair, *ame* ‘rain’ and *ame* ‘candy,’ is different between Tokyo and Kyoto-Osaka. In Tokyo, *ame* ‘rain’ is pronounced in the H* L pattern in contrast to *ame* ‘candy’ which is pronounced in the L H pattern. In Kyoto-Osaka, *ame* ‘rain’ is pronounced in the L H*+L pattern, in contrast to *ame* ‘candy’ which is pronounced in the H H pattern.

![Diagram](image.png)

**Figure 5.5.** Examples of bimoraic words in the Tokyo system and Kyoto-Osaka system (labeled as Osaka in the figure)

(Adapted from Fujisaki et al. 1976: 174; modified by the author)

The present section has highlighted contrasts in intonational systems between Tokyo and Kyoto-Osaka. As Kibe (2010) notes, differences in intonational patterns is one of the most commonly discussed dialectal features in Japanese public discourse due to the clear perceptibility of the distinction between East and West. By dividing the whole country into two while polarizing Tokyo and Kyoto-Osaka, intonation tends to appeal to people’s local identities, particularly in Kyoto and Osaka. Participants from Kyoto and Osaka in the present study commonly ridiculed people from Tokyo, commenting specifically on their intonation. For example, an Osaka
participant, Teru, during the interview, mimicked Tokyo speakers by uttering “choo kakkoiiyone [He is] super handsome” with the Tokyo intonational pattern. She then produced the same utterance with the Osaka pattern, and said, “I feel like ‘why the heck is the pronunciation like that?’ (nande koonaruno?tte naru, kotobaga).” Teru’s conversational partner Naho also commented on “people who do not have the correct [Kyoto-Osaka] intonation (chanto namattenai hito)” but try to speak the Osaka variety on TV, saying “I hate that! I feel like ‘the Osaka variety isn’t like that!’ (Are iya! Annan chaushi oosakaben ttenaru!).” To those from Kyoto and Osaka, intonation is a feature difficult to acquire by outsiders and is a marker of authenticity. The difficulty of acquiring intonation is supported by the above-mentioned study by Yoneda (1997). The significance to be noted at the end of this section is that unlike the features examined in Chapters 4 and 5, people are aware of the locality indexed by intonational patterns, which makes intonation more salient in construction of local identities.

5.3. Background II: Deaccented Negative Polar Questions

5.3.1. Deaccented Negative Polar Questions in Japanese

The intonational feature currently being investigated is the deaccentuation of NPQs. An example of a deaccented pattern of the utterance [kore] takaku nai? ‘Isn’t [this] expensive?’ shown earlier in the introduction of this chapter is repeated below as a reminder. The deaccentuation occurs in the negation form -nai? or -ne?, which is the monophthongized and shortened form of -nai. In conversational Japanese, there are two common ways to form polar questions. One is by using sentence-final particles that mark questions (e.g. -ka or -kai), and the other is simply through a rising intonation. The NPQs under investigation in this chapter occur only in the latter form since questions marked by final particles are considered less casual and are rarely
used in daily conversations by young people today. As noted below, the deaccentuated form is strongly associated with casualness and young people’s speech.

According to Tanaka (2010), the deaccented form emerged and rapidly spread out among the younger generation in the Tokyo area in the 1990s. In an early study (Tanaka 1993) on the feature, a survey demonstrated that the deaccented form was associated with young women, specifically “high school girls” and “gyaru”. As is noted in Chapter 2, gyaru refers to a young female fashion culture identified by its flashy style. However, more recent studies (e.g. Inoue 200, Tanaka 2010, Wakuda 2003) have shown that young men also use the feature with equal frequency. Tanaka (2010: 128) accounts for this change noting an increased use by men in the 2000s, during which time the feature established mainstream status in the area. Tanaka concluded that there is no particular gender indexicality found in the use (Tanaka 2010), but in fact, a survey in the same study by Tanaka still indicates a gendered meaning. The survey of high school students in the Tokyo area reveals the “urban”, “cute”, and “casual” image of the deaccented form, contrasting with the “correct” and “polite” image of the conventional form. The “cute (kawaii)” image implies that there is a feminine meaning of the feature, and the fact

Figure 5.1. (left) Conventional accented pattern

Figure 5.2. (right) New deaccented pattern

(Adapted from Hara and Kawahara 2012)
that the more masculine attribute “kakkoii ‘cool’” received a very low response (kakkoii received only 1.4% of the responses, while kawaii received 14.5% of the responses) strengthens the argument that there is a more feminine meaning associated with it.

The same survey study conducted by Tanaka (2010) reports a very different image, compared to the image of deaccented NPQs, toward the high rising terminal declarative (HRT) (shiri-agari), another linguistic feature commonly observed in young Japanese women’s speech (F. Inoue 2006). The HRT was already found in the late 1970s (Sibata 1977), and F. Inoue (2006: 215) notes that the feature, which he calls “the pseudo-questioning intonation,” began to appear widely in Japanese mass media around 1992 and became prevalent among young female speakers by 1994. While the HRT and the deaccented form emerged in mass media around the same time period, Tanaka’s (2010) survey study shows that the HRT has already fallen out of use because it received negative images such as “shallow (keisotsuna),” “obnoxious/abhorrant (nikurashii),” “uncool (kakkowarui),” and “child-like (kodomo)” in addition to the common images of the deaccented form such as “youth (wakamono)” and “casual (kudaketa).” Tanaka (2010: 147-48) argues that the negative image of the HRT in contrast with the positive image of the deaccented form is due to the fact that people noticed the HRT more quickly than the deaccented form because of its occurring at the phrase-final positions that are the most salient locus when marking pragmatic information in Japanese. While her previous study in 1992 observed frequent occurrences of the HRT among high school students, Tanaka (2010) notes that high school students today avoid the feature.

Since the deaccented form is predominantly a Tokyo feature, previous studies have discussed only the Tokyo variety. Tanaka (2010: 129) reports anecdotally on the use of the feature by a young male speaker from the Kinki region in a TV interview in the early 1990s, before the feature became enregistered. She notes, “In the mass media then, [the deaccented form] appeared infrequently, and no social discourse was attached [to deaccentuation].
However, considering that a young aspiring actor who was not residing in the Tokyo area was already using it, some stereotype peculiar to the entertainment industry might have begun to form [relating to the use of deaccentuation] around that time.” Though Tanaka does not further discuss this point, the link between the feature and the mass media is a relevant theme that will be discussed in the present study.

Many other previous studies of Japanese deaccented NPQs have focused mainly on the taxonomic distinctions based on production experiments and tended to rely on impressionistic observations to propose pragmatic functions of the feature. The most common underlying assumption is that it functions as an expression of seeking agreement (e.g. Tanaka 2010, Tsai 1996, Wakuda 2003). Tsai (1996: 45) hypothesizes that the deaccented NPQs sound less forceful to the addressee indicating from the beginning of the utterance that the phrase is an agreement-seeking expression. In a similar line of argument, Zawiszová (2012: 58) considers the use of the feature to be a sign of trying to avoid sounding too committed to the statement so as not to sound “too vulnerable” to the interlocutor’s negative response. According to Zawiszová, by indicating a lack of commitment, the speakers attempt to create a sense of “shared responsibility” with the addressees.

Hara and Kawahara’s (2012) study is exceptional in that the meaning of the deaccented form is investigated systematically based on a perceptual experiment. They first hypothesized the feature indicated evidentiality and tested the correlation between deaccentuation and availability of public evidence with an experiment. In the experiment, 14 participants were asked to judge the naturalness of different combinations of accentual patterns and degrees of public evidence availability on a 5-point scale. Results of the statistical analysis supported their hypothesis with a strong correlation between deaccentuation and available public evidence as well as a correlation between accentuation and unavailable public evidence.
In the following sections, semantic and pragmatic foundations of deaccentuation and NPQs will be reviewed. Previous literature on English and Japanese theoretically supports the impressionistic observations on the deaccented NPQs by Tsai (1996) and Zawiszová (2012) as well as the results of the experiment presented by Hara and Kawahara (2012). Though much of the foundational literature reviewed in the next section discusses only English, I assume that those theories are carried over into Japanese considering that those theories have been applied cross-linguistically, including in studies of Japanese. In their application to Japanese, some adjustments may be needed, which will be also noted in the next section. The key concept that serves as the basis of all the relevant literature reviewed in the remainder of Section 5 is the distinction between inside negation and outside negation proposed by Ladd (1981). I will argue that the Japanese deaccented NPQs are characterized by outside negation.

5.3.2. Meaning of Rising Intonation

As noted above, the NPQs under investigation in this chapter occur only in questions marked by a rising intonation and never in questions marked by sentence-final particles. Though I adopt the common practice of labeling these types of utterances “polar questions” to be in line with previous studies, it is in fact more precise to call them “rising declaratives” borrowing Gunlogson’s (2001) term. In her dissertation, Gunlogson (2001) investigated the meaning of rising intonation by comparing English declarative sentences with falling intonation (‘It’s raining’) and rising intonation (‘It’s raining?’). Since falling and rising declaratives constitute a minimal pair, differing solely in intonation, Gunlogson argues, any systematic differences between the two should be located in the contrast of the intonation. In addition to the relative “naturalness” of the rising declaratives as a question compared to the falling declaratives, Gunlogson implemented the falling declarative vs. rising declarative contrast in terms of
commitment vs. non-commitment. Based on the model, it is argued that the two types of declaratives express different levels of commitment to the propositional content of the declarative. This difference derives from the inherent contextual bias associated with declaratives, which constitutes a crucial contrast with interrogatives in general. By using rising intonation, according Gunlogson, speakers reconcile the bias with lack of commitment. While falling declaratives attribute commitment to the speakers, rising declaratives attribute it to the addressees.

### 5.3.3. Meaning of Negative Polar Questions

Implementing the notion of commitment is an innovative challenge put forth by Gunlogson (2001), but she is not the first scholar to have discussed biases conveyed by questions. One of the earliest studies that highlighted biased questions is Ladd (1981). Ladd (1981) proposed that NPQs in English should be subdivided into two categories based on a difference in scope of negation: “inside negation” vs. “outside negation”. To explain the distinctions between the two, Ladd provides the following examples:

**Example 1. Inside Negation**

(Situation: Bob is visiting Kathleen and Jeff in Chicago while attending the CLS)

Bob: I’d like to take you guys out to dinner while I’m here – we’d have time to go somewhere around here before the evening session tonight, don’t you think?

Kathleen: I guess, but there’s not really any place to go to in Hyde Park.

Bob: Oh really, isn’t there a vegetarian restaurant around here?

Kathleen: No, about all we can get is hamburgers and souvlaki.
Example 2. Outside Negation

(Situation: Kathleen and Jeff have just come from Chicago on the Greyhound bus to visit Bob in Ithaca)

Bob: You guys must be starving. You want to get something to eat?

Kathleen: Yeah, isn’t there a vegetarian restaurant around here – Moosewood, or something like that?

Bob: Gee, you’ve heard about Moosewood all the way out in Chicago, huh? OK, let’s go there.

According to Ladd (1981), Example 1 shows an “inside-negation” question, and Example 2 shows an “outside-negation” question. In Example 1, contrary to the speaker’s initial expectation, the speaker now infers that there might not be a restaurant and asks for a confirmation of this new inference. Here, the negation is inside the proposition in question because what is questioned is the inference. On the other hand, in Example 2, the speaker believes that there is a vegetarian restaurant and asks whether or not her belief is correct. In this case, the negation is outside the proposition in question, because what is questioned is the speaker’s belief, and not the proposition. The belief the speaker carries is “bias” in Gunlogson’s (2001) sense and called “positive epistemic implicature” by Romero and Han (2004). Romero and Han (2004) claim that proposed polar NPQs like Example 2 necessarily carry this implicature.

Ladd (1981) further observes that Positive Polarity Items (PPIs) and Negative Polarity Items (NPIs) disambiguate these two readings. That is, an NPQ which contains a PPI offers only the outside negation interpretation, whereas an NPQ with an NPI only carries the inside negation interpretation. Example 3 contains the PPI too, and Example 4 contains the NPI either (examples adapted from Romero and Han 2004).

Example 3. A: Ok, now that Stephen has come, we are all here. Let’s go!
Isn’t Jane coming too?

⇒ Implicature: Jane is probably coming.

Example 4. A: Pat is not coming. So we don’t have any phonologists in the program.

S: Isn’t Jane coming either?

⇒ Implicature: unexpectedly, Jane might not be coming.

Distinct cognitive statuses held by the biases in the two types of NPQs are highlighted by Reece (2007). Whereas inside-negation polar questions (INPQs) simply question an inference, outside-negation polar questions (ONPQs) assert a proposition. He demonstrates this point by testing whether or not some expressions that show illocutionary force can co-occur with the NPQs on the basis of Sadock’s (1974) diagnostics. For instance, sentence initial after all or yet can occur only with assertions, thereby, if it can be applied to an NPQ, the NPQ conveys an assertion. In Example 5, the bolded line presents both the inside-negation, indicated by including the PPI some, and the outside-negation, indicated by including the NPI any. This example shows that the reading of the outside-negation can be preceded by after all, but the inside-negation interpretation cannot. This leads Reece (2007) to conclude that the ONPQs convey assertion, while the INPQs are simply questions.

Example 5. A: Sue isn’t coming, so there’ll be no syntacticians there.

B: What do you mean?

After all, aren’t {some/#any} of the MIT syntacticians coming?

Since Ladd’s (1981) dichotomy of inside-negation and ONPQs is not enough to capture what kind of evidence is available for speakers’ inferences or beliefs, Büring and Gunlogson (2000)
propose the notion of contextual evidence. Contextual evidence is “[evidence] that has just become mutually available to the participants in the current discourse situation” (Büring and Gunlogson 2000: 7).

Example 6.  A: Where do you want to go for dinner? (no contextual evidence)
               B: #Isn’t there any vegetarian restaurant around here?
               C: Isn’t there some vegetarian restaurant around here?

Example 7.  A: I bet we can find any type of restaurant you can think of in this city.  
               Make your choice! (evidence for \(p’\))
               B: #Isn’t there any vegetarian restaurant around here?
               C: #Isn’t there some vegetarian restaurant around here?

Example 8.  A: Since you guys are vegetarians, we can’t go out in this town, where
               it’s all meat and potatoes. (evidence against \(p’\))
               B: Isn’t there any vegetarian restaurant around here?
               C: Isn’t there some vegetarian restaurant around here?

Based on Examples 6-8, Büring and Gunlogson concluded that there is compelling contextual evidence against the proposition in INPQs. Contrastingly, they concluded that, in the ONPQs, there is no compelling contextual evidence for the proposition, which means that there is either no contextual evidence or evidence against the proposition.
5.3.4. Meaning of Japanese Negative Polar Questions

As is mentioned earlier, conversational Japanese can make polar questions in two ways: One with question-marking sentence-final particles and the other solely with rising intonation. Borrowing Ladd’s (1981) notions of inside and outside negations as well as Büring and Gunlogson’s (2000) notion of contextual evidence, Sudo (2013) demonstrated that the differences in the phrase endings (whether or not a particle is used, or which particle is used) result in differences in the evidential and epistemic biases presented by the questions. Since the deaccented NPQs are reported only in the questions without phrase-final particles, discussions on NPQs with particles are put aside in this section. Examples 9-11 present examples of Japanese NPQs with different contextual evidence. A presents INPQs with the NPIs, and B presents ONPQs with the PPIs. Sudo’s (2013) conclusions are summarized in Table 5.4 below.

Example 9. (Context: I am in Osnabruck for the first time. My friend Daniel might or might not have been to this city before; no contextual evidence)

A: Doko-ka oishii resutoran sira nai?
    somewhere good restaurant know-NEG
    ‘Don’t you know some good restaurant?’

B: #Doko-mo oishii resutoran sira-nai?
anywhere good restaurant know-NEG
‘Don’t you know any good restaurant?’

Example 10. (Context: My friend Daniel has just said “there are all sorts of restaurants around here” while looking at a guidebook; evidence for $p$)

A: *Doko-ka* nihon-shoku *nai?*
   somewhere Japanese-food exist-NEG
   ‘Isn’t there some Japanese restaurant?’

B: #$Doko-mo* nihon-shoku *nai?*
   anywhere Japanese-food exist-NEG
   ‘Isn’t there any Japanese restaurant?’

Example 11. (Context: At a student meeting. Daniel is the student representative and knows who will be present today. Daniel has just said, “we are all here now. Shall we begin the meeting?”; evidence against $p$)

A: #$Dare-ka* hokani *ko-nai?*
   someone else coming-NEG
   ‘Isn’t someone else coming?’

B: *Dare-ka* hokani *ko-nai?*
   anyone else coming-NEG
   ‘Isn’t anyone else coming?’

(Sudo 2013: 14-15; modified by the author)
As is shown in Table 5.4, just like English examples presented in Ladd (1981), epistemic bias – bias toward a positive proposition – is implied only by ONPQ. For evidential bias, INPQs in Japanese contain contextual evidence against the proposition, which corresponds with the English INPQs. However, while there is no contextual evidence for the proposition in English ONPQs, there is no contextual evidence against the proposition in Japanese ONPQs. Rather, Japanese ONPQs imply contextual evidence for the proposition.

So far, the past few sections have reviewed previous studies of the two types of NPQs, called INPQs and ONPQs. These previous studies (Büring and Gunlogson 2000, Gunlogson 2001, Ladd 1981, Reece 2007, Romero and Han 2004, Sudo 2013) theoretically support the functions and meanings of the Japanese deaccented NPQs proposed by Hara and Kawahara (2012) and other studies, while revealing that the deaccented NPQs contain outside negation. Hara and Kawahara’s finding that the deaccented NPQs are more likely to be used when evidence to support the positive answer is available is in line with Sudo’s (2013) conclusion that Japanese ONPQs imply positive epistemic bias and contextual evidence for the proposition.
More impressionistic speculations by Tsai (1996) and Zawiszová (2012) suggested that deaccentuation functions to decrease the forcefulness of the utterance. This speculation underlines the assumption that ONPQs do assert propositions, which is argued by Reece (2007). Further, Tsai and Zawiszová’s speculation is supported by Gunlogson’s (2001) point that rising intonation indexes less commitment by the speaker, and the commitment is attributed to the addressees. While Japanese questions with conventional intonation do not exhibit any intonational cues for questions until the boundary tone, the deaccented questions provide signs for questions from the very beginning of the utterances. In sum, the deaccented form indexes a specific interactional stance where the speaker’s epistemic and evidential biases are asserted while the speaker expresses (or pretends as if there is) a lack of commitment to show a lack of vulnerability to disagreements, which constructs “urban cuteness.” The following sections further discuss intonational patterns and NPQs.

5.3.5. Tonal Compression of Japanese Negative Polar Questions

Intonational patterns are found to be relevant in distinguishing the meanings of the inside and outside negations expressed by Japanese rising declarative questions. Ito and Ohshima (2013) report that the ONPQs involve “tonal compression” while the INPQs do not. “Tonal compression” is defined as “weakening or tonal suppression of pitch movements due to lexical accents and/or phrase tones (but not due to intonation with such contours as question rises)” (Ito and Oshima 2013: 3)

---

29 Sicoli et al. (2015) found that speakers of a number of languages utilize initial pitch to differentiate questions designed to solicit agreement with an evaluation from questions which request information from the very beginning of the utterances. Analysis of pitch in 2,512 questions by 180 speakers in 10 different languages show that in agreement-seeking questions, pitch was divergent from a speaker’s median and was predictably in the top 10% of a speaker’s range.
Example 12. (Figure 5.6) \[ \text{Kore} \ warito \ amakunai? \quad \text{(warito is a near PPI)} \]
\[ \text{this} \quad \text{quite} \quad \text{sweet. Inf. Neg. Prs} \]
\[ \approx \text{‘Isn’t this quite sweet?’} \]

Example 13. (Figure 5.7) \[ \text{Kore} \ amari \ amakunai? \quad \text{(amari is a NPI)} \]
\[ \text{this} \quad \text{not very sweet. Inf. Neg. Prs} \]
\[ \approx \text{‘Is this not very sweet?’} \]

(Ito and Oshima 2013: 3; modified by the author)

Examples 12 and 13 both mean ‘isn’t this sweet?’ semantically, but Example 12 shows the outside negation being compatible with the PPI warito whereas Example 13 shows the inside negation with the NPI amari. Both figures exhibit the accented intonational pattern, with a difference in the tonal compression. Figure 5.6 presents only one AP with one pitch accent due to deletion of the accent in -nai, while Figure 5.7 presents two APs with the retained pitch accent in -nai.
While Sudo (2013) concluded that the inside negation polar questions do not connote any epistemic biases, Ito and Oshima (2013) argued that the inside negation polar questions could also index negative bias. Ito and Ohshima support this point by examining appropriate responses to the two types of questions with different accents.

Example 14.  Q: [AP ama’ku nai]? ‘Isn’t [it] sweet?’ (Outside Negation)

A1: Hai. ‘Yes.’

⇒ Indeed, it’s sweet

A2: Iie. ‘No.’

⇒ It is not sweet.
Example 15. Q: \([AP \ ama'ku] \ [AP \ na'i]\)? ‘Isn’t [it] sweet?’ (Inside Negation)

A1: Hai. ‘Yes.’

⇒ Indeed, it’s not sweet.

A2: Iie. ‘No.’

⇒ It is sweet.

(Ito and Ohshima 2013: 4-7; modified by the author)

While both the questions amakunai? ‘Isn’t [it] sweet?’ in Examples 14 and 15 function to seek for agreement, the responses clearly show that the propositions implied by the question are differently interpreted by the interlocutors. It is apparent that the question in Example 14 successfully communicates the speaker’s bias/belief that it is sweet in contrast with the question in Example 15, which communicates the speaker’s opposite bias/belief that it is not sweet.

5.3.6. Meaning of Tonal Reduction

As is mentioned in an earlier section, there is an effect of downstep, by which the pitch height of an AP is reduced when following another accented AP in an IP. Apart from this downstep effect, post-focal reduction is commonly discussed in previous studies (Deguchi and Kitagawa 2002, Ishihara 2004, Kitagawa 2005, Sugahara 2003, Tomioka 2009). Figure 5.8 illustrates post-focal reduction in the utterance kyo‘nen ro’omani ikima’shita ‘[I] went to ROME last year’. The word ro’oma ‘Rome’ is the focus of the utterance, indicated with more defined prominence than the pitch accent of the most left AP, and the pitch register of the following AP ikima’shita is compressed. Post-focal reduction, according to Venditti, Maekawa, and Beckman (2008: 487), works against the occurrence of any further pitch accents within an IP after a word that bears the focus of the utterance.
In the discussion of the tonal reduction, Ito and Oshima (2013) offer an explanation based on the post-focal reduction. Their explanation is that in the ONPQs, –nai is tonally compressed since the negation is “part of ground”, but in the INPQs, the negation is “part of the focus”, so –nai is not tonally compressed (Ito and Oshima 2013: 5). As Hara and Kawaraha (2012) and Ishihara (2003: 32, 74) point out, the tonal reduction is not compatible with deaccentuation in that whereas the deaccentuation refers to a complete deletion of the lexical pitch accent, post-focal reduction refers to the Fo lowering of the prominence while retaining the pitch accent. It is also true that the discourse functions of the two phenomena are different (Hara and Kawahara 2012). However, Ito and Oshima’s (2013) study is still significant in that it provides an example where the particular intonational pattern – reduction of Fo contour – indexes biased meaning of NPQs in Japanese.

5.3.7. Meaning of Deaccentuation
It has also been shown in several studies of information structure (e.g. Allerton 1978, Baumann and Riester 2013, Halliday 1967, Schwarzschild 1999) that information status of ‘newness’ and ‘givenness’ plays an important role in intonational prominence. New information is usually accented, but given information triggers deaccentuation, where the pitch accent of the element is deleted. According to Halliday (1967: 206), “new” is defined as textually and situationally non-derivable information but also as “contrary to some predicated or stated alternative,” whereas “given” is clearly and consistently defined as anaphorically recoverable by reference, substitution, or ellipsis. In the same line of argument, Chafe (1976: 31) notes that “[t]he principle linguistic effects of the given-new distinction in English, and perhaps all languages, reduce to the fact that given information is conveyed in a weaker and more attenuated manner than new.” Though there is no study that has investigated specifically the association between the informational status of givenness and the Japanese deaccented NPQs, it is reasonable to claim that the deaccentuation helps a speaker frame the proposition as a “given fact” on the basis of the epistemic and evidential biases.

5.4. Current Study

5.4.1. Introduction

While the previous section has highlighted meanings of the deaccented NPQs, particularly in Tokyo Japanese, this section examines how the feature is actually used by non-Tokyo speakers as well as how it is used by Tokyo speakers. The data of the present study shows that all the Tokyo participants except for one use the deaccented form, particularly with adjectival predicates. More surprisingly, the data reveals that five Kyoto participants and one Osaka participant also use the same form. Locally specific meanings of Japanese deaccented NPQs in construction of femininities will be further explored through discourse analysis and supported
with ethnographic information in the discussion sections. While the meaning of “urban cuteness” discussed in the previous section is applicable in the use of the deaccented form in Tokyo, an additional meaning is proposed in Kyoto and Osaka, namely Tokyo-centric gendered ideology, which revolves around “urban cuteness.” In this ideology, an “ideal” young female figure is described as urban, trendy, stylish, style-conscious, and cute like the “typical” Tokyo girl.

5.4.2 Procedure

For the analysis of the deaccented intonational pattern, NPQs in the form of –nai? or –ne?, the shortened form of –nai?, were coded within 60-minute segments of conversations for each participant. All the regional negative forms commonly found in Kyoto and Osaka such as –chau? and –hen? were not included in the data since the initial H pitch of those regional forms prevents further intonational rise and, therefore, deaccentuation is not possible with them. In total, 188 tokens from Tokyo, 122 tokens from Kyoto, and 153 tokens from Osaka were collected.

Together with whether or not the question is accented, the grammatical categories of the predicates were coded to examine linguistic constraints. In the data, six grammatical categories were identified: i-adjectives (iA), na-adjectives (nA), verbs (V), nouns (N), nominalized verbs (Vn), and verbs inflected with the pattern of i-adjectives (Vi). Japanese adjectives are grouped into either i-adjectives (e.g. kawaii ‘cute’, omoshiroi ‘funny’) or na-adjectives (e.g. kireina ‘pretty’, henna ‘weird’) based on inflectional patterns. For the first four categories (i-adjectives, na-adjectives, verbs, and nouns), only the non-past negative inflectional forms were coded since deaccentuation does not apply to the past forms. Nominalization of Japanese verbs can be realized by attaching the nominalizer –n or –no to the dictionary forms of verbs. For example, the verb iku ‘go’ is nominalized into iku-n or iku-no. The –no form is more formal and old-fashioned (Tsukuba Language Group 1992), so only the form –n was found in the present data.
The nominalized verbs inflect just like regular nouns, therefore, only the non-past negative forms were coded for this category. Finally, an innovative and “ungrammatical” inflectional pattern for verbs was found in the data. Many participants simply attached the non-past negative ending of *i*-adjectives, *-ku nai*, to the dictionary form of verbs to make negative verbs. For example, the conventional inflectional rule makes the verb *iku ‘go’* into *ika-nai ‘not go’*), but the innovative pattern makes it *iku-ku nai ‘not go’*. While the conventional rules classify Japanese verbs into three groups and inflect verbs accordingly based on the groups of the verbs, the innovative pattern simplifies the rule by proposing the sole inflectional pattern: dictionary form + *ku nai*. Note that this innovative form is not commonly seen as a negative declarative form; but it is a phenomenon found in the context of NPQs.30

Mixed effects logistic regression was conducted with R, including speaker as a random effect. The dependent variable was accentual pattern (accented or deaccented), and the independent variables were region and word category (wordcat).

5.4.3. Results

30 However, some researchers (F. Inoue 1988, Ishii 2011, Kitamoto1995) report that this innovative form is commonly used in the verb *chigau ‘differ’* among young speakers in the Tokyo area. F. Inoue (1988) argues that the use of this particular verb in the form of an *i*-adjective is due to the fact that the meaning of the verb (‘differ’) concerns the quality of an object that is usually described by adjectives, rather than an action that is usually referred to by verbs. In spite of the prevalence of the form, Ishii (2011) shows that over 40% of the 50 respondents from the same population expressed discomfort about the declarative use of the form. On the other hand, as for the use of the form in NPQs, 49 out of the 50 respondents answered that the form is comfortably acceptable. This also supports commonly observed use of the innovative form in the present study.
Table 5.5. Mixed effects for model of deaccentuation

Table 5.5 summarizes the results, and Figure 5.9 presents their distributions in each region with the number of occurrences. The intercept in intonational pattern was accented, and the intercept in region was Kyoto. The intercept in word category was i-adjective. In the results, region presents three-way distinctions. Tokyo exhibits strikingly more deaccentuation than Kyoto, while Kyoto exhibits more deaccentuation than Osaka, which thus indicates that Tokyo exhibits more deaccentuation than Osaka. These patterns are illustrated in Figure 5.9 as well.
For the word categories, adjectives and nouns are likely to be deaccented while verbs are not likely to be deaccented. Interestingly, plain verbs and nominalized verbs are less likely to be deaccented compared to i-adjectives, but verbs that inflect just like i-adjectives are more likely to be deaccented.

5.4.4. Discussion

5.4.4.1. Grammatical Categories
The finding that adjectives and nouns are more likely to be deaccented than verbs can be explained in terms of the nature of the lexical items under those categories. As is discussed at length in earlier sections, the deaccented NPQs carry implicatures of biases and beliefs toward the positive propositions. The biases, particularly epistemic biases, are constructed by the speakers’ subjective perspectives. Subjectivities are more compatibly expressed by adjectives than verbs as is clear from the common label “subjective adjectives.” Those subjective adjectives by definition underline the speakers’ subjective views, which succinctly represents the speaker’s biases. In fact, all of the deaccented adjectives observed in my data were subjective adjectives such as sugoi ‘great,’ kawaii ‘cute,’ kireina ‘pretty,’ yabai ‘crazy,’ kimo ‘creepy,’ and henna ‘weird.’ Also, many of the deaccented nouns in the data were used in a similar fashion to the adjectives. For example, by looking at a picture of a male model in very tight jeans wearing some makeup in a male fashion magazine, a Tokyo participant said, “kore onnaja nai? ‘Isn’t this a woman?’” In this utterance, the speaker did not really think the model was a woman; rather, what she meant was that the model looked feminine. Here, the noun represents certain qualities that adjectives usually refer to.

Compared to adjectives and nouns, verbs in NPQs require more specific conditions to communicate speakers’ biases. One-time action verbs usually do not imply the speakers’ biases. For example, if a wife is wondering when her husband is leaving home, she may say, “Don’t [you] go yet? (mada ika nai?).” The verb ‘to go (iku)’ does not convey her bias toward the positive proposition and thus cannot be deaccented. Verbs can imply biases when they show habitual acts or continuing states resulted from previous actions. For example, questions such as “Isn’t [he] always eating? (itsumo tabete nai?)” (habitual) and “Aren’t [you] annoyed? (mukatsuka nai?)” (continuing state) imply the speakers’ biases. In fact, the second example (“Aren’t [you] annoyed?”) came from my data, and the verb ‘to get annoyed (mukatsuku)’ was
deaccented. Other deaccented verbs found in my data are *iru* ‘to exist’, *niteru* ‘to look alike’, *shitteru* ‘to know’, and *kimatteru* ‘to be decided/set’ (transitive verb). 

Among the verb categories, the regression model has also found that the verbs inflecting like *i*-adjectives are more likely to be deaccented. This suggests that the prosodic pattern of the deaccented form of *i*-adjectives is highly established, and it even makes verbs deaccented. In particular, the deaccentuation of the verb *chigau* ‘to differ’ in the innovative form *chigau-ku nai? ‘Doesn’t it differ?’ was commonly found. As F. Inoue (1988) argues, the use of this particular verb in the inflectional form of the *i*-adjective is prevalent among young speakers since the meaning of the verb (‘to differ’) concerns the quality of an object that is usually described by adjectives rather than verbs.

### 5.4.4.2. Pattern in Tokyo

Unsurprisingly, the results have shown that deaccentuation occurs in Tokyo more than Kyoto and Osaka. The deaccented form was consistently used by all the Tokyo participants, especially with adjectives, except for one speaker. Before highlighting this exceptional Tokyo speaker, I will discuss the use of the deaccented NQPs in gender and regional identity construction in Tokyo. Among previous studies of the Japanese deaccented NQPs, the most widely agreed function of the feature is agreement seeking. Expressions seeking agreement such as tag questions have been considered a sign of a lack of confidence and weakness by early scholars of language and gender (e.g. Lakoff 1973, Holmes 1982, 1984, 1995)\(^{31}\), but the deaccented NQPs in fact imply confidence with the epistemic and evidential biases by framing the utterance as if it

---

\(^{31}\) It should be noted that other studies (e.g. Algero 1988, Cameron, McAlinden, and O’Leary 1989, Winefield, Chandler and Bassett 1989) point out that there are many other functions served by tag questions, not all of which work to subordinate women. Algero (1998) reports that tag questions can be used in a rather aggressive manner in British English, and Winefield, Chandler and Bassett (1989) note that tag questions are employed by a female psychotherapy patient to show independence.
were a given fact. While implying confidence, the speakers show (or pretend as if there is) a lack of commitment to the proposition so that they appear less vulnerable to possible disagreement. This stance corresponds to the “cold” and “distant” image of the Tokyo variety presented in Figure 1.4 in Chapter 1. Considering that cuteness has often been associated with weakness, intimacy, and a lack of confidence in previous literature (e.g. Kinsella 1995, Miller 2004, Moon 2010, Kajino 2011) it is interesting to note that this feature is perceived as indexing “urban cuteness” by the high school students in the Tokyo area (Tanaka 2010). Recall that the Tokyo participants did not mention “cuteness” as an image of their own variety, but the Kyoto and Osaka participants commonly perceived the Tokyo variety spoken by females as “cute” (see the detailed discussion in Chapter 1).

While all three regions consist of metropolitan areas, “urbanness” is the most prominent trait of Tokyo. Tokyo is, in fact, a larger city than Kyoto and Osaka, but also the participants’ views on the regions support this point. In the interviews, I asked the participants to tell me about the aspects of their own region that they like. A majority of the Tokyo participants mentioned the convenience of the area because of the subway systems and availability of “everything (nandemo),” and a few talked about Tokyo’s trendiness. In contrast, all of the Kyoto participants used the words “peaceful (heiwa),” “nature (shizen),” or “country (inaka).” Those who described Kyoto as “country” compared Kyoto with Tokyo and Osaka. For example, Sakura noted that “[Kyoto] is neither as noisy (gayagaya shiteru) as Osaka nor as sophisticated (senren sareteru) as Tokyo.” In Osaka, two participants mentioned the convenient subway systems and described Osaka as “urban (tokai),” but the rest of the participants brought up friendly and funny Osaka people. These participants’ descriptions of their own regions show that “urbanness” is a conspicuous characteristic of the Tokyo identity for the Tokyo participants. The Kyoto participants recognize Tokyo’s “urbanness” as a contrastive trait with their region on which the identity of Kyoto partly relies, while it is not the case in Osaka.
Despite the fact that previous studies as well as the present study have demonstrated that the deaccented form is a mainstream feature in the Tokyo area, a Tokyo participant Takako never used the feature during the recoding. Takako, a freshman at the time of the fieldwork, was recorded with her classmate Yasuko. Takako and Yasuko were the only participants from Tokyo who did not wear their hair dyed or permed. In the recorded conversation, they mainly talked about the fashion styles of girls on the campus, part time jobs, and types of men they would marry in the future. Yasuko expressed a negative evaluative stance toward girls who spend a lot of time and energy on their looks and clearly stated, “I don’t have any interest in fashion.” Takako was, on the other hand, surprised by how stylish girls on the campus were and was afraid to be seen as unfashionable. She even hoped that someone would take her shopping and transform her into someone fashionable, though she also said, “I can’t help thinking that if I had a lot of money to spend on fashion, I would buy better things like a computer. With the computer, I can watch more anime, play more games, and read more books.”

In their recording, Yasuko exclusively used the deaccented form when the predicates are adjectives, just like the other Tokyo speakers. For instance, in Excerpt 5.1, Yasuko criticizes girls who wear short skirts by using a deaccented NPQ.

Excerpt 5.1. (Takako and Yasuko are talking about girls’ fashion styles)
1. Yasuko: Zettaitowa iihirenaikedosa, ashi hutoiyatsuni kagittesa,  
   ‘Though I can’t say it’s always the case, but among girls with fat legs,  
2. minisuka-ritsu takaku nai? (Deacc)  
   isn’t the possibility of [wearing] short skirts high?’  
3. Takako: Un, takaiyone!  
   Yes, it’s high!  
‘(laugh) Since those legs are fat, I think like ‘why don’t you hide them?’’

5. Takako: *Tamani hosokute sugoku kawaii ko irukedosa...*

‘Sometimes there are skinny and really cute girls...

6. *Ashi hosokute suggoku niautteno kuyashii (giggle)*

‘I feel jealous of them being good-looking with the skinny legs (giggle).’

In Line 2, Yasuko makes fun of girls who expose their “fat legs” by saying, “(Among girls with fat legs,) isn’t the possibility of [wearing] short skirts high? (*minisuka-ritsu takakunai?*)” with deaccentuation. In response to this NPQ, Takako agrees with the positive proposition in Line 3. Takako’s response shows that Yasuko has successfully communicated the bias to Takako. It is also noteworthy that after once agreeing with Yasuko’s criticism in Line 3, Takako brings up an opposing statement in Lines 5-6. This shows that Takako has avoided directly conflicting with the clearly stated belief of Yasuko in Line 2.

In contrast to Yasuko, Takako never used the deaccented form in the recording. Excerpt 5.2 presents an example of interactions where Takako uses an NPQ.

---

**Excerpt 5.2. (Takako is looking at pictures of jewelry in a female fashion magazine)**

1. Takako: *Teyuukasa, nankasa, kooyuuno nande tsuketenno minna.*

‘By the way, why do people wear something like this?’

2. *Omota’ku nai?* (Acc)

‘Isn’t [it] heavy?’

3. Yasuko: ... *Chiechaino dattara.. sokomade omokuwa naidarookedo...*

‘.. If it’s small... then I guess it wouldn’t be that heavy...’
4. Takako: *Teyuuka sokomade hito mitekurennokanato omoundayone.*

‘I mean, I wonder if other people look at the small details (of your fashion).’

In Excerpt 5.2, Takako talks about jewelry presented in a female fashion magazine and uses an NPQ as in “why do people wear something like this? Isn’t [it] heavy?” The utterance ‘*Omotaku nai?* Isn’t [it] heavy?’ in Line 2 was articulated with the conventional accent. In response to the NPQ, Yasuko, after a pause, answered that she thinks that it would not be very heavy if the jewelry were small. From this response, it is clear that Yasuko decided that the NPQ is a simple question that solicits information, though the pause in turn suggests indeterminacy. However, Takako’s following utterance (“I mean, I wonder if other people look at the small details (of your fashion”) reveals the fact that Takako actually had bias toward the positive proposition with the belief that since people would not notice such small details like the jewelry shown in the magazine, the weight of the jewelry is too heavy and not worth experiencing. Obviously, Takako failed to communicate the bias to Yasuko, and this miscommunication was caused by the failure to use the deaccented form by Takako.

The use and absence of the deaccented NPQs by Yasuko and Takako are symbolic of their stances toward girls’ fashion and the dominant gendered ideology in the heterosexual marketplace (Eckert 1996). This dominant gendered ideology describes a typical or mainstream young Tokyo female as an urban, trendy, stylish, style-conscious, and cute girl just like those presented in the fashion magazines. Yasuko takes a clear stance against the mainstream gendered ideology in the heterosexual marketplace displayed in the female fashion magazines and styles of girls on the streets of Tokyo by stating that she does not have any interest in fashion and by criticizing them. By taking this stance, Yasuko treats the mainstream young female figure as something attainable, just as the feature of “urban cuteness” (i.e. the deaccented form) was to her. On the other hand, although Takako expresses admiration toward the girls
who successfully embody the mainstream fashion styles, she assumes the figure is something
difficult to attain, just as the deaccented form is to her.

5.4.4.3. Pattern in Kyoto

The finding that Kyoto exhibits quite a few occurrences of the deaccented form is surprising
since finding any Kyoto and Osaka speakers using this predominantly Tokyo feature is
unexpected. As the map of Japanese regional intonational patterns (Figure 5.6) has shown,
Kyoto and Osaka are categorized under the same intonational system, which is distinguished
from the Tokyo intonational system. Because of the clear contrast between the Tokyo system
and the Kyoto-Osaka system, intonation is a feature that is most commonly discussed as a
regional marker by the general public. Also as a previous study (Yoneda 1997) shows, acquiring
a new intonational system is a more difficult task than acquiring other phonetic features, which
leads intonation to be an authenticity marker. Being situated in this cultural context, the use of
the Tokyo feature by non-Tokyo speakers is salient. For the Kyoto and Osaka speakers to
produce this deaccented form, there are two steps to go through. Since deaccentuation does not
occur in the Kyoto-Osaka local intonational patterns, they have to first adapt a Tokyo accentual
pattern, and then deaccent it. Adopting a Tokyo accentual pattern is perceived as inauthentic,
and deaccenting manifests a deliberate gender performance that embodies the young Tokyo
female figure, which represents “urban cuteness.”

In the data of this study, the Kyoto participants use the deaccented form significantly
more than the Osaka participants. A close examination among the Kyoto speakers reveals that
deaccentuation is used exclusively by five speakers, and the other 10 speakers never use the
form. The five speakers who use deaccentuation are Akari, Mitsuki, Rika, Sayumi, and Mika.
Akari and Mitsuki were recorded together, and Rika and Sayumi were recorded together. Mika was recorded with her friend Eri, who never uses the deaccented form. Common among these five speakers is that they are all freshmen and that they all graduated from all girls’ high schools. Further, they are all big fans of mass media such as manga, anime, TV dramas, and boy idols, and their talk in the recorded conversations centers on those media. It is also revealed in their conversations that all of them embrace the heterosexual ideology offered by the media where female figures are represented with Tokyo speech. The heterosexual ideology describes an “ideal” young female figure as a “typical” young Tokyo girl who is urban, trendy, stylish, style-conscious, and cute. What follows in this section shows excerpts of their conversations with some examples of their use of the deaccented form.

**Mika**

The conversation between Mika and Eri provides interesting data to observe miscommunication between a user and a non-user of the deaccented form. While both Mika and Eri are athletes on college teams (Mika is a long-distance runner, and Eri is a volleyball player), their recorded conversation clearly shows the different interests of the two. Mika enjoys girls’ mangas in which emotions and romantic relationships of teenaged girls are a strong focus, but Eri has no interest in the genre. Mika, during the interview for this study, told me that she might be unconsciously using women’s language when “[her] heart is melting (*tokimeku*)” from reading girls’ mangas. Excerpt 5.3 shows Mika’s use of the deaccented form. Preceding this excerpt, Mika was telling Eri how good girls' mangas are, where Eri dismissed it by saying, “After all, it's just about girls’ crushes.”

---

The fact that these speakers were conversational partners may suggest that they are accommodating to each other’s speech. As I will note in Chapter 7, it would be interesting to investigate the conversations as dyadic and how much each participant’s speech works to shape the other’s in a future study.
Excerpt 5.3. (Mika has just explained to Eri how good girls’ romantic mangas are)

1. Eri: *Paradaisu-kisumo yondakoto naishina.*
   ‘I haven’t even read *Paradise Kiss.*’

2. Mika: *Ah, uchimo nai.*
   ‘Ah, I haven’t either.’

3. Eri: *Nainkaina.*
   ‘You haven’t either?’

4. Mika: *(laugh) E, naikedosa, e, demo mi-*mitaku ne?* *(Deacc)*
   ‘(laugh) Well, I haven’t, but **don’t you want to watch [the movie]**?’

5. Eri: .. *U-um?*
   ‘.. U-um?’

6. Mika: *Mi- demo meccha mite mitai.*
   ‘Mi- but I really want to watch it.’

7. Eri: *Iran.*
   ‘Not me.’

Excerpt 5.3 shows miscommunication between Mika and Eri due to Mika’s use of deaccentuation. After talking about one girls’ manga, Mika brings up the movie created based on the manga by using the deaccented NPQ *mitaku ne?* ‘Don’t [you] want to watch [the movie]?’ in Line 4. Note that the Japanese desiderative ending –*tai* converts the verb *miru* (‘to watch’) into an *i*-adjective as in *mitai*. Mika constructs the NPQ with –*ne*, the shortened form of the negation suffix –*nai*, as in *mitaku ne?* However, Eri does not comprehend this utterance of the deaccented form as Line 5 indicates. In Line 6, Mika initially attempts to repeat the NPQ (*mitaku ne?) but instead, decides to explicitly state the point as in “but I really want to watch it”.

188
These few lines of the interaction between Mika and Eri reveal that Mika attempted to communicate the positive proposition through the deaccented NPQ but the form was unexpected to Eri. Mika was also aware that Eri was not familiar with the form, and therefore, she elaborated the underlined meaning of the deaccented NPQ to avoid further miscommunication.

**Akari and Mitsuki**

Akari and Mitsuki both love anime and videogames. During the recorded conversation, they imitated the speech of both male and female characters of anime and TV shows several times. Especially, Akari repeatedly performed her favorite line said by one male TV show character that she loved. In the interviews with me, Akari said that she thinks the Tokyo variety is easy to understand, and Mitsuki said, “When I’m watching anime, I get affected by the ways the characters talk.” In Excerpt 5.4, Akari is showing a picture of Mitsuki doing ‘cosplay’. Cosplay, the abbreviation of ‘costume play’, refers to the popular practice among anime, manga, and video-game fans in which the participants dress like specific characters and enjoy role play. In the picture, Akari is showing Mitsuki a picture of herself dressed just like a male anime character and posed with another female cosplayer who dressed as a female character.

Excerpt 5.4. (Akari is showing Mitsuki a picture of Mitsuki’s cosplay from a year ago)

   *(giggle) Don’t [you] feel nostalgic? This picture.*

   *(I do feel nostalgic. I remember I was doing this.)*
In Line 1 in Excerpt 5.4, Akari uses the deaccented NPQ Natsukashiku nai? ‘Don’t [you] feel nostalgic?’ with higher pitch. In Japanese, natsukashii ‘feeling nostalgic’ is an i-adjective. Unlike Mika and Eri’s interaction in Excerpt 5.3, the meaning of the deaccented NPQ is successfully communicated to Mitsuki, and she agrees with the positive proposition “I do feel nostalgic.” In this interaction, they frame the event from just a year ago as if it were from a long time ago, and by doing so they talk as if they were presently more mature and experienced. By using higher pitch, Akari positions herself as a woman who is mature enough to have dramatic experiences just like female characters in TV shows.

After Excerpt 5.4, while looking at the same picture, Mitsuki situates herself in a hypothetical relationship by saying, “If these two [in the picture] were a real couple, they would be a punk couple.” During the recording, Mitsuki enjoyed men’s fashion magazines more than women’s fashion magazines, because (as she claimed) she identifies her style more with men’s style. Regarding women’s fashion, she commented that it is a subject she is not good at (nigatena). In the interview, she told me that she used conventional masculine self-referential forms such as boku and ore, and even more unconventional masculine form ore-sama (‘respectable I’), which is used to index over-self-confidence and narcissism in a joking manner, when she is “excited (koofun shiteru)”. Though Mitsuki seems to orient more toward male gender in the virtual world, she still holds heterosexual ideology as well as a strong interest in dating men. Preceding Excerpt 5.5, Mitsuki tells Akari that the ideal first date with a man would be going to ride a roller coaster.

Excerpt 5.5. (Mitsuki is explaining to Akari why she wants to ride a roller coaster on a hypothetical first date)

1. Mitsuki: Demo itsumo heeonyakaratte kanjiwa naikana.

   ‘But I guess, because my life has always been calm/boring.’
2. Akari:  
   *Ah.. (giggle)*

   ‘Ah.. (giggle)

3. Mitsuki:  
   *Datte.. koi made heeonyattara iyaja nai?* (Deacc)

   ‘Because.. **Wouldn’t [it] be terrible** if even my relationship were boring?’

In the explanation as to why she wants to ride a roller coaster on a first date, Mitsuki uses the
deaccented form of the NPQ as in “**Wouldn’t [it] be terrible** if even my relationship were boring?”
Her heterosexual interest is clearly expressed in this utterance with the deaccented form.
Elsewhere, when looking at pictures of wedding dresses, Mitsuki also tells Akari about the kind
of wedding dresses she wants to wear in the future. At the same time, she expresses her concern
about whether she can find a partner and expresses regret that she is not a male since she would
not have had problems finding a female partner if she were a man. These utterances by Mitsuki
lead to the conclusion that her orientation toward the male gender in hypothetical relationships
is reflective of her lack of confidence in participating in the heterosexual marketplace as a female
because of the homogeneous female figure portrayed in the mass media. The use of
deaccentuation and the narcissistic masculine pronoun indicates her confidence in participating
in the heterosexual ideology in hypothetical relationships and the virtual world of cosplay;
however, those features were never used in her speech about real-life concerns such as shopping
for her clothes and finding a boyfriend or future husband.

**Rika and Sayumi**

In their recorded conversation, Rika and Sayumi almost exclusively talk about TV celebrities
including idols, young actors, musicians, and fashion models. During the interview with me,
Rika and Sayumi both associated the Tokyo language variety with mass media. To Rika, unlike
the majority of the participants from Kyoto, hearing the Tokyo variety does not make her uncomfortable “because it’s normal in TV shows (terebi mitetara hutsuudakara iwakan nai),” and Sayumi sometimes mixes up her local variety with the Tokyo variety since “[she] watches too much TV (terebi misugite hyoojungo ga mazaru).” What makes their conversations different from the other participants’ who also talked about TV celebrities is that Rika and Sayumi spent more time talking about the topic than any other participants and take the issues more seriously by positioning themselves as experts. Sayumi identifies herself as a serious fan of a very popular boy band and tells Rika that the recent scandal of a member of the group being involved with a famous actress has severely broken her heart. In some parts, Rika and Sayumi even become competitive toward each other over expert knowledge on those celebrities and the TV industry in general. In Excerpt 5.6, Rika is showing Sayumi a picture of a female fashion model’s (who usually wears a lot of makeup) face without makeup. In the interaction, they both use deaccented NPQs and agree with each other that the model looks pretty without makeup on.

Excerpt 5.6. (Rika is about to show Sayumi a picture of a female model on her cell phone)

1. Rika:  Atta. **Kawaii** **nai?** *(Deacc)*
   ‘Here it is. **Isn’t [she] cute?**’

2. Sayumi:  **E, mecha kawaii.**
   ‘Oh she looks very cute.’

3. Rika:  **Kore, Komori Jun-yade.**
   ‘This is Komori Jun.’

4. Sayumi:  **... Kawaii. Meiku shinaihooga yoku nai?** *(Deacc)*
   ‘... She looks cute. **Doesn’t [she] look better** without makeup on?’

5. Rika:  **Omou!**
   ‘I agree!’
In the first line of Excerpt 5.6, Rika uses the deaccented form and expresses her belief that the model looks pretty in the picture. Sayumi agrees with Rika in Line 2 and reconfirms the agreement by using the deaccented NPQ in Line 4. While the deaccented form was used to indicate an agreement in Excerpt 5.6, the next excerpt shows strong disagreement in the conventional local accent.

Excerpt 5.7. (Rika is pointing at a picture of Sayumi’s favorite male idol in a magazine)

1. Rika:  
   \textit{Paama niawanaiyone}.  
   ‘He doesn’t look good with the perm, does he?’

2. Sayumi:  
   \textit{E! Niatteru’ku nai?} (Acc)  
   ‘What! Doesn’t [he] look good?’

In Excerpt 5.7, Sayumi’s strong objection to Rika’s claim that the male idol does not look good is expressed by the rapid pitch fall of the conventional accent despite the fact that the underlined statement expressed by the NPQ is the positive proposition (i.e. ‘He does look good’). Here, Sayumi uses the unconventional inflectional pattern by attaching the ending of $i$-adjectives $–ku nai$ to the verb niatteru ‘to look good.’ Her surprise and slight anger can be heard with a rapid increase of intensity and pitch in the beginning of Line 2. Throughout the conversation between Rika and Sayumi, the deaccented form is never used when they express conflicting views. As is demonstrated by Hara and Kawahara (2012), the deaccented form presupposes that the common ground supports that there is some shared evidence for the positive proposition, and therefore, the form cannot express conflicting views. Additionally, it should be highlighted that Sayumi’s NPQ in Line 2 was accented with the Kyoto-Osaka regional intonational pattern. The
verb *niatteru* ‘to look good’ was pronounced with the initial H pitch, which can be found only in the Kyoto-Osaka intonational system, and not in the Tokyo system (see Tables 5.1 and 5.2 for the different accentual patterns in Tokyo and Kyoto-Osaka). I argue that Sayumi’s choice to use the local accentual form in this context shows that the local form is more suited for the expression of impulsive emotional reactions like strong objections or slight anger than the Tokyo pattern. As is discussed in Chapter 1 (Figure 1.4), the “fake” image of the Tokyo variety is consistent among the Kyoto and Osaka participants, and the deaccented form is part of the conscious gender performance for the Kyoto speakers. Neither “fakeness” nor “urban cuteness” conveys Sayumi’s strong objections or slight anger, but her authentic accent pattern does.

In this section so far, I have argued how Kyoto participants’ fandom and their deep involvement in media products are linked to their use of the deaccented form. Here, it should be emphasized that the speakers’ participation in the heterosexual gendered ideology presented by the mass media is the significant key rather than simply the consumption of the media. This point is made clear in the comparison of the Kyoto participants featured earlier in this section and Orie, another anime fan, who is a senior student from Kyoto. Orie’s life is surrounded by her favorite anime characters, such as stuffed dolls of *One Piece* characters attached to her backpack (see Figure 5.10) and a picture of a female character printed on her *Evangelion* wallet (see Figure 5.11). Her daily routine includes staying up all night to watch minor animes aired on TV. Her passion toward anime is so serious that she dreams of becoming a voice actress, but financial circumstances prevent her from going to a voice acting school and moving toward the dream. During the interview, she performed voices of a few characters from the popular anime series *Evangelion* and *The Melancholy of Haruhi Suzumiya* with the exact lines memorized.

---

33 In the interview with me, one Osaka participant, Miki, noted that she could not link the Tokyo variety with strong emotions like anger, saying, “When Tokyo people are angry, I feel like ‘Are you really angry?’” Another Osaka participant, Midori, also said that her non-Osakan friend finds the Osaka variety intimidating since it makes the Osaka speakers sound angry.
While Orie is very involved in one genre of the mass media, her speech does not display the same pattern with other media fans in Kyoto. Other anime fans in Kyoto like Akari and Mitsuki, adopt some linguistic features of anime characters such as high F0, wide F0 range, and faster speech rate as well as the deaccented form in their speech while embracing the heterosexual ideology presented in the media. In contrast, Orie draws a clear boundary between her speech and voice acting. In her acting voice performed during the interview, the pitch, pitch range, speech rate, and intensity of her speech increased, and she became more articulate in general, which presented a sharp contrast with her “normal” speech. She also commented that she uses Japanese Women’s Language (JWL) when she “gets into [anime] character (yakuni narikiru).” While Akari and Mitsuki entertain fantasized thoughts of virtual relationships with the media products, Orie never talks about anime characters as virtual romantic partners in her conversation. Her lack of active participation in romantic matters may be rooted in past experiences of being bullied by male classmates. This leads me to speculate that to Orie the virtual world of anime is not a space to acquire heterosexual practices; instead it might be a
sanctuary from the harsh real world dominated by the hegemonic heterosexual ideology, which creates hierarchy among women.

### 5.4.4.4. Pattern in Osaka

Throughout the Osaka conversation data, only one occurrence of deaccentuation by Midori was found. From this single occurrence, it is difficult to determine the meaning of the variable in Osaka, but one point I can make from this example is that just like the speakers in Kyoto featured above, Midori is also a freshman that is very into a young male idol. In her conversation with Kei, who is more interested in comedies than the idol scene, Midori attempted to obtain an agreement on how good-looking the male idol is several times, by asking Kei “Kakkoyoku nai? ‘Isn’t [he] handsome?’” with the conventional (Osaka) accent.

The following excerpt shows the one occurrence of deaccented NPQ use by Midori.

Excerpt 8. (Midori and Kei are talking about a concern of going bald)

1. Kei:  *Shanpuutoka shitetarasa, kekkoo nukennenyanka.*
   ‘When I do shampoo and stuff, a good amount of hair falls out.’

2. Midori:  *Ahh.. e, demo nukeru jikitte aruku nai?* *(Deacc)*
   ‘Ahh.. well, but **isn’t there** a season when more hair falls out?’

   ‘I wonder if there is.’

The example of the deaccented form used by Midori above does not present strong evidence for meanings that are peculiar to Osaka. In Line 2, Midori uses the negative question to comfort her
friend who is worried about losing hair. A possible reason why she uses the deaccented form here is to display her assurance toward the proposition that Kei’s concern is only a temporal issue, which could make Kei feel better. In this case, Midori’s use of the form indicates her understanding of its discourse function, which is likely learned through the mass media, but it does not indicate her participation in the Tokyo-centric virtual romance unlike the Kyoto participants’ use of the form. In fact, what makes Midori different from the five Kyoto users of the deaccented form is that Midori situates her romantic desire not only in the virtual media world, but also in the real-life heterosexual marketplace. In the conversation, Midori tells Kei about an older man she met at a party who tried to hit on her. While Midori expresses disinterest in the man by making fun of him, she is also intent on highlighting his value in the marketplace based on his seriousness toward her as well as his economic success. Midori notes that the man paid for her to take a taxi home even though it was a long distance.

Other than Midori, no participant from Osaka used the deaccented form. Compared to the Kyoto participants, the Osaka participants exhibit a more consistent pattern. One proposed reason of this consistency among the Osaka participants is that the language variety of Osaka is more established in terms of daily linguistic use than the Kyoto variety. Though the Kyoto variety holds a highly enregistered image as is presented in Figure 1.5 in Chapter 1, the Kyoto participants’ subjective perspectives on their own language use do not correspond with the stereotypical “classy” and “soft” image of the Kyoto variety. As a result, 12 out of the 15 Kyoto participants responded negatively to the interview question, “Do you think you speak the Kyoto variety?” All the 12 Kyoto participants who do not think that they speak the Kyoto variety claim that they speak the Kansai (‘Western’) dialect instead. Among the other three participants, two responded that they probably speak the Kyoto variety because other people had told them that they do, and the other rejected answering the question since she simply did not know what the Kyoto variety was in terms of actual speech. On the other hand, to the same interview question
“Do you think you speak the Osaka variety?” all of the 16 Osaka participants responded positively. This clearly indicates that the Osaka speakers carry ideas about what is considered as the Osaka speech and identify themselves with the ideas. I claim that the somewhat rigid shared ideas restrict the Osaka speakers from using linguistic features considered inauthentic, particularly the Tokyo intonational pattern because of its saliency. In this regard, the Osaka speakers are more constrained in their stylistic practices than the Kyoto speakers.

5.5. Conclusion

This chapter has investigated the meaning of deaccented NPQs in construction of gender and regional identities across Tokyo, Kyoto, and Osaka. In the first half of the chapter, Japanese prosody was also reviewed while highlighting the clear contrasts between the Tokyo system and the Kyoto-Osaka system. The general public acknowledges the contrasts between the two systems, and intonation is considered to be a marker of authentic locality particularly for the Kyoto and Osaka speakers.

Previous literature of the target feature, the deaccented NPQs, has widely documented that the feature is an expression of agreement seeking dominantly used by young speakers from the Tokyo area. It has also been reported that the deaccented form is considered as “urban” and “cute” among high school students in the Tokyo area (Tanaka 2010). Building on previous literature, this chapter has first explored the pragmatic meanings and functions of the deaccented NPQs. Ladds’ (1981) notion of inside- and outside-negations and its related research provided supports for the epistemic and evidential biases expressed by the deaccented form.

The latter half of the chapter has discussed the present study. Statistical analysis was conducted to examine the use of the deaccented form and the accented form across the three regions and across the grammatical categories of the deaccented predicates. Interestingly, the
regions presented three-way distinctions. Tokyo exhibits more deaccentuation than Kyoto, while Kyoto exhibits more deaccentuation than Osaka, thus Tokyo exhibits more deaccentuation than Osaka. For the grammatical categories, adjectives and nouns are likely to be deaccented while verbs are not likely to be deaccented.

To further investigate the meaning of the deaccented NPQs in the construction of gender and regional identities, discourse by some “exceptional” speakers from Tokyo, Kyoto, and Osaka were analyzed in detail. The close examination of interactions by Takako, the only Tokyo speaker who never uses the deaccentuation, and Yasuko revealed that the use and absence of the deaccented form is symbolic of their differing stances toward the mainstream young female figure who is trendy and cute in the heterosexual marketplace. On the other hand, analysis of the use of NPQs by the five Kyoto speakers highlighted that the use of the deaccented form indicates their alignment with the (Tokyo-centric) heterosexual ideology presented in the mass media. Because there was only one occurrence of the deaccented form found in the Osaka data, the local meaning of the form remained unclear. On the other hand, the consistent use of the local intonational form by the Osaka participants is indicative of their highly enregistered local language variety.
CHAPTER 6
Style and Meaning: The Recipe for Femininities

6.1. Introduction

While Chapters 3-5 have each concentrated on a different linguistic feature, this chapter will explore meanings of breathy voice and various realization of /s/ by locating the variants in the participants’ stylistic practices. Interspeaker variation in individual styles will shed light on intra-regional diversity and ways of negotiating the regional stereotypes and constructing femininities. Examinations of individual participants’ linguistic and non-linguistic styles reveal that higher degrees of breathiness are observed in the speech of the participants who display mature femininity. The analysis also shows how breathy voice with high pitch or low pitch constructs two different types of feminine figures – the sensitive and traditional mature women and the strong and bold mature women. Furthermore, the analysis demonstrates the roles that the sound quality of sharpness of the fricative /s/ plays in construction of femininities. The quality of sharpness links to the harsh Osaka image, but this chapter will also showcase how some Osaka participants disalign themselves from this stereotypical image by avoiding high COG of /s/.

6.2. Approach to Style

Various definitions of style have been proposed in sociolinguistics literature. In earlier variation literature (e.g. Labov 1972), style was conceptualized as a unidimensional continuum between casual and careful speech across different degrees of attention paid to speech. However, a more recently developed theory presents a multidimensional model that takes into account various
factors including the formality of the speech situation, the social relationship with the audience and with wider groups (e.g. ethnic groups, gender groups), and speakers’ internal goals of projecting particular types of stances, personal characteristics, and personas in the interactions (Schilling 2013a), and thereby proposes a more sophisticated approach to style, and by extension, to how language styles shape and are shaped by identity considerations. Under this approach, style is defined as “a clustering of linguistic resources, and an association of that clustering with social meaning” (Eckert 2001: 123). Styles are also considered “the product of ideology, insofar as they are posited by speakers ... as more or less clearly defined and socially specifiable collections of coinciding symbolic forms bound to particular social groups via metapragmatic stereotypes34” (Bucholtz 2009: 146-47). Additionally, Irvine (2001) calls attention to the “distinctiveness” of style and the process of stylistic differentiation, because a particular style cannot be described independently from others.

The present study adapts this multidimensional approach to style, aligning with the third wave approach to style and social meaning (Eckert 2012; see Chapter 1 for an overview of waves of sociolinguistic variation study). Following Eckert (1996, 2000), Bucholtz (1999), and Johnstone (1995), I consider features of linguistic style to cluster not only with other co-occurring linguistic features but also with non-linguistic features of stylistic practices, since language is interconnected with style in action and appearance (Eckert 1996: 184). Material style plays a central role in identity formation especially among young people in late capitalist societies like the United States and Japan (Bucholtz 2007), and therefore, is integrally related to and provides important information for studying linguistic style (Eckert 2008: 457). Sclafani (2009: 627-28) also emphasizes that analyzing linguistic aspects of style in isolation does not provide sufficient evidence to uncover the social meaning of a given linguistic feature, and social

34 Metapragmatic stereotypes are defined as “culture-internal models of utterance indexicality associated with speech variants” (Agha 2007: 148).
meaning can be derived only through consideration of both linguistic and non-linguistic aspects of style.

Regarding the third wave variationist approach to style, as is emphasized by Schilling (2013a), we should acknowledge that under such an approach, there might be some potential risk of neglecting important aspects of stylistic variation. First, by having a strong focus on speakers’ agentivity and creativity, third wave studies sometimes overlook the dilemma that speakers can never be fully agentive and creative because following some structures and norms is necessary for a speaker to effectively convey meaning (Schilling 2013a: 342). No matter how hard one attempts to construct a unique and innovative speech style, she or he inevitably reuses pre-existing linguistic resources which already carry associations with particular social groups and/or character types through previous repetitive use. Social forces that constrain people’s creativity and agentivity are particularly evident in the construction of gender identities (Schilling 2011). In investigations of Japanese female speech, we need to consider the socially powerful force of JWL (Japanese Women’s Language), which determines ‘appropriate’ forms and functions of language Japanese women should use.

At the same time, Schilling (2013a) further claims that the focus on agentivity and creativity may also lead us to wrongly assume that every single component of style is a result of self-conscious placement for the deliberate act of identity creation. Instead, as Bucholtz and Hall (2004) emphasize by using the term *intersubjectivity* rather than identity, identities can be constructed as an unintended relational phenomenon without full intentionality by, for instance, habitual practices. What Agha (2007) discusses as *the relational self* also highlights the mutual aspect of identity construction. According to Agha (2007: 238), the relational self is “a figure of identity that depends on the ability of persons to assign personae to each other by attending to matters of *appearance* and *conduct*” (italics by the author). Co-constructed identities are a key concept of my approach to style, and neglecting this can cause misinterpretation of individual
speakers’ linguistic practices. Furthermore, not all linguistic resources carry stylistic meanings, and even among the ones that do carry stylistic meanings, some features and meanings are more or less salient.

To avoid such possible misinterpretations, Schilling (2013a: 343) encourages mixed-methods research utilizing both quantitative and qualitative sociolinguistic approaches (including both discourse analytic and ethnographic approaches) to help us best understand what patterns of variation mean to the speakers who use them. For example, in her study of language variation across ethnic groups in Robeson County, NC (Schilling-Estes 2004), she analyzes how phonological and morphosyntactic variables are used in unfolding sociolinguistic interview discourse between a Lumbee Indian and an African American to understand the meanings underlying patterns found by large-scale quantitative investigation. In particular, while the big-picture patterning of r-lessness in Robeson County was not clear, the analysis of the interview interaction uncovered multiple meanings indexed by the same feature, such as African American ethnicity, Southern locality, and vernacularity and in addition revealed the dynamic nature of identity construction, which is “a product of unfolding talk” (190). In Chapter 5 of the present dissertation, I also investigated how the deaccented form of negative questions in Japanese is used in unfolding conversation by the five Kyoto speakers who use deaccentuation to determine the locally specific meaning of the feature because the quantitative patterns across the three regions suggested that the only meaning was Tokyo regionality.

While previous analysis chapters, especially Chapters 3 and 4, have focused on the big-picture patterning of breathy voice and fricatives, this chapter is concerned with the placements of those features along with co-occurring linguistic and non-linguistic features in styles, and further explores the social meaning of the variables based on the individual styles. By doing so, this chapter considers construction of styles as the process of both entextualization and recontextualization (Bauman and Briggs 1990). Entextualization is a process by which texts are
generated by extracting discourse out of its original interactional context and incorporating elements of its history of use (Bauman and Briggs 1990: 73). Once the texts are produced through entextualization, the texts become subjects of recontextualization through social practices that do not simply correlate preexisting elements of the texts but rather reshape the texts through the new situational contexts (Bauman and Briggs 1990: 76). For example, in the study of staged comedic performances of the Osaka Auntie speech style and Tokyo Hostess speech style (Kajino 2014), one of my findings was that the Osaka Auntie speech exhibits a significantly faster speech rate than the Tokyo Hostess speech. While fast speech rate contributes to replication of the Osaka Auntie stereotype with the ready-made meanings through previous usages that have entextualized the figure, its associated linguistic characteristics, and the social meanings associated with those features (such as carelessness and roughness; see Section 6.3.2 below for the meaning of speech rate), the variant acquires a new association by being used in the Osaka Auntie speech performance in a exaggerated form with Osaka regionality or Osaka Auntie-ness. Through the processes of entextualization and recontextualization, linguistic features are not simply resources used to create styles by providing pieces of preexisting meanings; rather, linguistic features can take on new meanings by being placed in a certain style.36

Finally, while style in sociolinguistic variation has been widely understood in terms of intraspeaker variation, this chapter will concentrate on interspeaker variation in individual styles. As is clear from Rickford and Eckert’s (2001: 1) statement that the three principal components of sociolinguistic variationist studies are “linguistic or internal constraints, social or inter-speaker constraints, and stylistic or intra-speaker constraints,” research on interspeaker

35 While Tokyo Hostess is also a middle-aged woman like Osaka Auntie, Tokyo Hostess exhibits sharp contrasts with Osaka Auntie in that she is a speaker of JWL and values traditional womanhood.

36 Bauman (1996: 302) also notes, “one of the key issues on which understanding of the process must rest is the dynamic tension between the ready-made, socially given element, that is, the persistent cultural entity that is available for recontextualization in performance, and the emergent element, the transformation of this entity in the performance process.”
variation has studied the relation between variation and social parameters such as class, gender, ethnicity, social networks, identity, local categories, and ideology through cross-group comparisons, while studies of intraspeaker variation (e.g. Blom and Gumperz 1972, Coupland 1981, Bell 1984, Rickford and McNair-Knox 1994, Schilling-Estes 1998, 2004, Rickford and Price 2013) have focused on style-shifting of individual speakers in accordance with various factors including the formality of the speech situation (i.e. attention to speech; see Labov 1972), social relationship with the audience (i.e. audience design; see Bell 1984), and speakers’ creative use of linguistic resources to project particular types of stances, personal characteristics, and personas in interactions (i.e. speaker design; see Coupland 2007, Schilling 2013a).37 Hernández-Campoy and Cutillas-Espinosa (2012: 6) summarize the focus on the social component through interspeaker examinations as “deterministic and system-oriented (language as a collective system: langue)” and the focus on the stylistic component and intraspeaker examinations as “more social constructivist and speaker-oriented (language as individual performance: parole),” but this chapter examines individual performance through interspeaker variation. By focusing on variation across speakers in each of the three regions, my goal is to shed light on intra-regional and interpersonal diversity and scrutinize different ways to negotiate the local language varieties and local stereotypes as well as different ways to construct femininities.

6.3. Linguistic Features of Investigation

37 While social and stylistic components are two different components, they have been investigated concurrently since the inception of variationist sociolinguistics, since patterns of variation across styles lends invaluable insight into the social evaluation of linguistic features and therefore also reasons for observed patterns of variation across social classes and, importantly, over time in the course of ongoing linguistic changes (Labov 1966; see also Schilling-Estes 2004). Also, the synchronized relation between intra-speaker and inter-speaker variation is highlighted in the notion of the Style Axiom proposed by Bell (1984, 2001), which states that “[v]ariation on the style dimension within the speech of a single speaker derive from and echoes the variation which exists between speakers on the ‘social’ dimension” (1984:151).
6.3.1. Main Features

In this chapter, I focus on H1-H2 as a measure of breathiness (measured for the vowel [a]), and COG of /s/ as a measure of sharpness in the analysis of style. As is discussed in Chapter 3, the formula H1-H2 indicates the difference between the increased amplitude of the first harmonic of a vowel and the amplitude of the second harmonic and corresponds to the open quotient of the glottal waveform – a ratio of the time period in which the glottal airflow increases and decreases to the total period of vocal fold oscillation. A larger H1-H2 value is a salient acoustic property of breathy voice. COG of a fricative indicates the frequency around which the greatest concentration of frication energy is observed, and a higher value indicates fronted articulation of the fricative. The COG values presented in this chapter are converted into ERB (Equivalent Rectangular Bandwidth) on a perceptual scale. This chapter examines mean H1-H2 and COG values of each participant for analyses of individual participants’ uses of these features. The mean values were calculated with R using the measurements obtained in Chapters 3 and 4.

Among the various metrics examined in Chapters 3 and 4, these two metrics (H1-H2 and COG of /s/) are the foci of this chapter since these presented cross-regional variation more clearly than other metrics, and also these are both the most widely used measures for breathy voice and realizations of fricatives in previous research (see the background sections in Chapters 3 and 4). For the same reason, the fricative /s/, and not /ɕ/, is highlighted since /s/ is the most extensively studied fricative in terms of gender.

6.3.2. Additional Features

For multidimensional analysis of individual participants’ stylistic practices, I examine more global (IP-level) features such as F0, F0 range, speech rate, and phrase final lengthening, in addition to H1-H2 and center of gravity (COG) of /s/. These additional features are chosen for
investigation in order to include a longer domain of linguistic practice in addition to the analyses of short time windows. Additionally, previous literature has reported associations between these features and gender, which makes the features salient in speakers’ stylistic practices. Also by attending to the IP (Intonational Phrase)-level analyses of individual speech, the analysis of these additional features presents a more global view of stylistic practices, while H1-H2 and COG are analyzed at the phoneme-level as is shown in Chapters 3 and 4.

**Fo and Fo Range**

As is discussed in Chapter 3 (Section 3.4.2.1), F0 has been widely examined in its association with gender. Especially in Japanese, high F0 is closely associated with femininity, and studies by Van Bezooijen’s (1995) and Yuasa (2011) show that Japanese women utilize higher F0 levels than Dutch and American women. Van Bezooijen also found that high pitch in female speech is perceived as more attractive by Japanese listeners, while Dutch listeners perceive medium and low pitch to be more attractive. Higher F0 carries an ideological link to Japanese womanhood since it is considered as a feature of JWL. In Podesva’s (2007) study on intra-speaker stylistic variation, a gay professional, Heath, uses a wider pitch range along with higher pitch levels in casual settings than in a professional settings in order to construct an expressive and flamboyant “diva” persona (defined as a person who is characterized by beauty and off-putting superiority (Podesva 2007:292)). Heath even utilizes falsetto and creaky phonations to expand his F0 range to the physical extreme. Conversely, Camp (2009) reported that the Japanese lesbian speakers in her study exhibited a narrower pitch range along with a lower mean pitch than the heterosexual speakers. The perception experiment in the same study showed that a narrow pitch range and a low pitch are both perceived to index masculinity and lack of emotionality. Yuasa (2008) also argues that a wider pitch range indicates the speaker's emotional excitement and
demonstrates how Japanese speakers restrict pitch range when speaking with people with whom they are not very familiar since Japanese cultural norms limit the expression of such emotions in non-intimate relationships. However, Ogino and Hong’s (1991) analyses of utterances produced by professional voice actors and by non-actors shows that in the case of female speakers, wider F0 range is perceived as more polite, though the correlation was not found for male speakers. Intimacy is a stereotypical quality of women’s collaborative interactional styles (e.g. Tannen 1990, Holmes 1993), and this intimate indexical meaning of a wide pitch range further leads to the ideological connection that exists between certain types of feminine personas and wide pitch ranges.

**Speech Rate**

Conflicting results have been reported in previous studies of gender effects on speech rate production. Some studies (Block and Killen 1996, Kowal, O’Connell, and Sabin 1975, Walker 1988; Ryalls et al. 1994, Quéné 2008, Van Borsel and De Maesschalck 2008) found no difference between males and females in speech rate, while others showed a higher speech rate in males (Binnenpoorte et al. 2005, Fitzsimons, Sheahan and Staunton 2001, Yuan, Liberman and Cieri 2006, Verhoeven, De Pauw and Kloots 2004). No studies have reported a higher speech rate in females, which suggests a possible association between female speech and slower speech rate.

Though there is not a clear correlation between speech rate and gender categories found in previous studies, the relationship between speech rate and listener perception of personal attributes discussed in a number of perception studies provides clues to draw some characteristics of the speakers. Miller et al. (1976) found that speakers with faster speech rate were perceived to be more persuasive, knowledgeable, intelligent, and objective than slower
speakers. Smith et al. (1975) demonstrated that the faster rates were perceived as being indicative of greater competence. However, it was also shown by Smith et al. (1975) that both the faster and slower rates were viewed as less benevolent than a “normal” rate. Apple, Streeter, and Krauss (1979) examined the influence of speech rate manipulations on judgments of persuasiveness, fluency, emphaticness, and nervousness and found that the slowest speakers received the lowest ratings on the variables whereas the normal, not the fastest, rates received the highest ratings. Therefore, it is concluded by the authors that decreasing rate had a “deleterious effect on a speaker’s perceived persuasiveness, fluency, and emphaticness” (723), but an increasing rate did not necessarily enhance the perception of those characteristics.

In Japanese, Ogino and Hong’s (1991) analyses of utterances produced by professional voice actors and by non-actors found that utterances produced with slower speech rate and longer pauses are likely to be perceived as careful and polite (teineina) in both female and male speakers. Hirose et al. (1997) reported that a careless and rough (zonzaina) speaking style shows a faster speech rate than neutral speech. Ofuka et al. (2000) studied the effect of speech rate on people’s perception of politeness and concluded that the utterances sound polite when speech rate is “moderate” but sound impolite when speech rate is increasingly faster or slower. Laver (1994) speculates that this finding by Ofuka et al. is a “universal tendency,” since fast speech may express irritation or urgency, while a slower rate may indicate hesitancy, doubt, or boredom in statements, or sympathy or encouragement in questions and commands (Ramsaran 1989; cited in Laver 1994).

**Final Lengthening**

Vowels at the end of a domain (utterance, intonational phrase, prosodic word) tend to be longer than corresponding non-final vowels. Final lengthening has been found in many languages all
over the world, including Japanese (Beckman and Pierrehumbert 1986, Kaiki et al. 1992, Ueyama 1996), English (Oller 1973, Umeda 1975, Wightman et al. 1992, Aylett 1999), German (Delattre 1966:190), Mandarin (Ho 1977:449, Fon and Johnson 2004), Cuban Spanish (Oller 1979: 334), Hebrew (Berkovits 1991:62), French (Fletcher 1991), Taiwanese (Peng 1997: 382), Hungarian (Hockey and Fagyal 1999: 315), Italian (Van Santen and D’Omeperio 1999) and Dutch (Rietveld et al. 2004). Elongated phrasal endings have also been noted in the visual modality of American Sign Languages (Stokoe 1978: 31) and finger spelling (Coulter 1993: 266). One explanation of the phonetic final lengthening is due to a general deceleration of articulation in coarticulatory anticipation of the static position the speakers will hold during pause (Klatt 1976: 1212; Edwards et al. 1991). In other words, when one goes from rapid movement, as in speech, to standstill, as in pause, it is easier to make the transition gradually. Kim (1974) states that intensity and Fo, along with subglottal air pressure, tend to decrease toward the end, and in order to saliently make up for this loss, lengthening occurs. Cooper (1976) claims that final lengthening is communicatively important because it gives the speaker a bit more time to prepare the next utterance.

Though sociolinguistic research of final lengthening is sparse, final lengthening is commonly identified with females in a few studies of East Asian languages. Kajino’s (2010) study on stylistic variation of the Japanese pop queen Hamasaki Ayumi showed that the use of a high degree of final lengthening with low pitch and narrow pitch range at talk shows helps her construct an immature and rebellious public image, because of the underlying notion that final lengthening is associated with childlikeness and is hence inappropriate for adults, especially, in public spheres. In the analysis of prosodic rhythm among three female characters and three male characters in a Chinese TV drama series, Callier (2011) revealed that the degree of final lengthening is robustly greater among female characters than male characters. Moon’s (2014) perceptual experiment on the phrase-final intonation LHL% in Korean revealed that degrees of
final lengthening positively correlate with the listeners’ ratings on childlikeness, a manipulated femininity (‘aegyo’), and whininess. In a similar vein with Moon’s finding, one of the Tokyo participants of the present study noted final-lengthening (“gobio nobasu”) as a stereotypical feature of the burikko speech (ultrafemininity presented through bogus innocence (Miller 2004); see the discussion on burikko in Chapter 1).

6.3.3. Measuring Procedures of the Additional Features

For the measurement of participants’ mean F0 levels and F0 ranges in the present study, their speech was parsed into IPs using PRAAT (Boersma and Weenink 2013). To include speech across different topics, 25 IPs were collected from each 20-minute segment of their conversational recording and 100 IPs per speaker were obtained. Recall that each conversation recording included three sessions on different topics (women’s fashion, men’s fashion, and careers and jobs) which each lasted 20 to 30 minutes, following a 10 minute-long warm-up. Within each IP, the PRAAT script noted the F0 level every 0.005 seconds and calculated the mean F0. Also the F0 ranges were calculated by subtracting the minimum F0 from the maximum F0 within IPs. All the IPs with measurement failures/errors were manually deleted, and the number of the IPs coded was 4,473 (1,451 from Tokyo, 1,474 from Kyoto, and 1,548 from Osaka).

To include various topics, I coded speech rate in 25 IPs every 20 minutes and obtained 100 IPs per speaker. For each IP, the final mora was first segmented from the rest of the utterance to eliminate the effect of final lengthening, and the number of moras per second was calculated (See Chapter 5 for the basic timing of Japanese). For example, in the utterance, “konnande iinenkedo-na (‘something like this would be fine’),” the duration from the beginning of the IP to the second to the last mora (“konnande iinenkedo”) was measured as 1.1 second. Since
“konnande iinenkedo-” contains 11 moras, the speech rate of the utterance was calculated as 10 moras/sec. IPs with the following characteristics were excluded from the data: those with less than five moras; those ending with long vowels (due to the impossibility of segmenting the final single mora); those containing longer vowels for emphasis and/or emotionality; and those showing interrupted fluency due to such causes as hedges, repairs, laughs, or overlaps. For the Osaka and Kyoto data, one-mora words (e.g. me ‘eye’ and se ‘height’) that were produced with elongated vowels were counted as two moras (e.g. mee ‘eye’ and see ‘height’) since monomoraic nouns of Tokyo Japanese tend to be realized with prosodic lengthening as two moras in Kyoto and Osaka (Labrune 2012: 253). Also segments where the participants are reading the text of magazines or directly quoting someone else were also excluded. After manually removing all measurement errors/failures, 4,111 IPs (1,335 from Tokyo, 1,351 from Kyoto, and 1,425 from Osaka) were coded.

Finally, degree of final lengthening was calculated for each IP as the ratio of the duration of the final mora to the average duration per mora. In the example “konnande iinenkedo-na (‘something like this would be fine’),” the average duration per mora of “konnande iinenkedo-” was 0.1 second. The duration of the final mora “-na” was measured as 0.2 second, so the score of final lengthening of this utterance is 2.\footnote{38 Though I did not control the segmental content in considering speech rate and final lengthening, it should be noted that segmental content can be a significant factor because different vowels, by their nature, vary in duration. In a future study, analysis of speech rate and final lengthening should consider this fact.}

212
Figure 6.1 summarizes individual participants’ mean values of H1-H2 and COG of /s/. The Tokyo participants’ values are displayed as crosses, the Kyoto participants’ displayed as circles, and the Osaka participants’ are presented as triangles. From the figure, it is noticeable that the Osaka participants are clustered together while the Tokyo and Kyoto speakers exhibit more

6.4. Intra-Regional Diversity

Figure 6.1 summarizes individual participants’ mean values of H1-H2 and COG of /s/. The Tokyo participants’ values are displayed as crosses, the Kyoto participants’ displayed as circles, and the Osaka participants’ are presented as triangles. From the figure, it is noticeable that the Osaka participants are clustered together while the Tokyo and Kyoto speakers exhibit more
diversity in the COG of /s/ and H1-H2 means. The consistency among the Osaka participants is indicative of the highly enregistered language variety. As is discussed in Chapter 1, the image the participants hold toward the Osaka variety was concrete and consistent across the different regions, and the image also conforms to the Osaka participants’ perception of their own speech. This finding shows that people, including Osaka speakers, have a clear idea about how the Osaka variety actually sounds and should sound, which makes Osaka speakers more constrained in their linguistic use. In addition to breathy voice and /s/, we saw a consistent pattern among the Osaka participants in the use of deaccentuation of negative polar questions in Chapter 5.

The Kyoto variety, which presents diverse patterns, as shown in Figure 6.1, in contrast, is only ideologically constructed and not enregistered as an actual audible speech variety. Although the variety carries a highly established image (“classy” and “calm” geisha image) as we saw in Chapter 1, most of the Kyoto participants do not think that the variety they are speaking is “the Kyoto variety.” Instead, they claim that they are speaking the Kansai (‘Western’) dialect. One Kyoto participant said that she had no idea what Kyoto speech actually sounds like. This uncertainty allows Kyoto speakers to exhibit more diverse linguistic features compared to the Osaka speakers. Finally, the Tokyo variety is consistently enregistered as hyōjungo (‘the standard language’), but these particular measurements (i.e. H1-H2 and COG of /s/), with the diverse values, do not exhibit consistency and “standardness.” This also supports the obvious claim that the “standardness” is merely ideological and not a reality.

6.5. Summary of Intra-Regional Variation

Table 6.1 below summarizes the 46 participants’ mean values of the six linguistic features, H1-H2, COG of /s/, F0, F0 range, speech rate, and final lengthening. Parentheses indicate ranks of the values among all participants; the top 10 rankings and the bottom 10 rankings are bolded for
each feature.

In the next section, I will investigate variation across individual speakers within each region to examine how language styles shape and are shaped by identity considerations. I will also explore the meanings of different degrees of breathy voice and the different levels of COG of /s/, drawing clues from linguistic and non-linguistic aspects of styles. Following Irvine (2001), individual styles are investigated by comparison with contrastive style since styles gain meaning only in relation to other styles. In what follows, I will first highlight breathy voice by examining variation in individual mean H1-H2 values. The examination shows that participants who display mature femininities use higher degrees of breathy voice. Then, I will focus on frequency energy of /s/ by looking at variation in individual mean COG values of /s/. The analysis will showcase the roles the sharp /s/ plays in construction of femininities.
## Table 6.1. Summary of individual means and ranks across linguistic features

<table>
<thead>
<tr>
<th>Region_no.</th>
<th>Speaker</th>
<th>H1-H2 (dB)</th>
<th>/s/ COG (erb)</th>
<th>F0 (Hz)</th>
<th>F0 Range (Hz)</th>
<th>Speech Rate (mora/sec)</th>
<th>Final Lengthening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokyo_1</td>
<td>Aika</td>
<td>-1.408 (19)</td>
<td>17.715 (44)</td>
<td>233.868 (30)</td>
<td>170.39 (11)</td>
<td>10.284 (17)</td>
<td>1.61 (16)</td>
</tr>
<tr>
<td>Tokyo_2</td>
<td>Akiko</td>
<td>-3.214 (27)</td>
<td>25.204 (18)</td>
<td>231.071 (33)</td>
<td>122.11 (36)</td>
<td>8.853 (38)</td>
<td>1.206 (44)</td>
</tr>
<tr>
<td>Tokyo_3</td>
<td>Yasuko</td>
<td>-8.224 (43)</td>
<td>20.341 (40)</td>
<td>269.53 (8)</td>
<td>175.95 (8)</td>
<td>9.093 (35)</td>
<td>1.881 (5)</td>
</tr>
<tr>
<td>Tokyo_4</td>
<td>Chami</td>
<td>-5.896 (38)</td>
<td>22.54 (34)</td>
<td>249.333 (21)</td>
<td>154.96 (19)</td>
<td>9.154 (34)</td>
<td>1.65 (13)</td>
</tr>
<tr>
<td>Tokyo_5</td>
<td>Fuyu</td>
<td>-11.03 (45)</td>
<td>20.424 (39)</td>
<td>297.396 (2)</td>
<td>202.78 (5)</td>
<td>8.275 (43)</td>
<td>1.392 (31)</td>
</tr>
<tr>
<td>Tokyo_6</td>
<td>Hana</td>
<td>0.04 (11)</td>
<td>18.74 (41)</td>
<td>335.674 (29)</td>
<td>96.02 (45)</td>
<td>10.002 (22)</td>
<td>1.523 (23)</td>
</tr>
<tr>
<td>Tokyo_7</td>
<td>Takako</td>
<td>0.004 (12)</td>
<td>26.61 (12)</td>
<td>195.411 (46)</td>
<td>113.29 (39)</td>
<td>11.212 (5)</td>
<td>1.741 (7)</td>
</tr>
<tr>
<td>Tokyo_8</td>
<td>Jun</td>
<td>2.591 (7)</td>
<td>20.991 (37)</td>
<td>199.26 (44)</td>
<td>210.69 (4)</td>
<td>7.748 (46)</td>
<td>2.467 (1)</td>
</tr>
<tr>
<td>Tokyo_9</td>
<td>Mei</td>
<td>-6.785 (42)</td>
<td>17.907 (43)</td>
<td>286.073 (4)</td>
<td>212.84 (34)</td>
<td>9.641 (27)</td>
<td>1.22 (41)</td>
</tr>
<tr>
<td>Tokyo_10</td>
<td>Megu</td>
<td>-2.06 (22)</td>
<td>25.082 (21)</td>
<td>246.204 (22)</td>
<td>129.33 (31)</td>
<td>8.134 (44)</td>
<td>1.485 (29)</td>
</tr>
<tr>
<td>Tokyo_11</td>
<td>Ruka</td>
<td>-2.06 (23)</td>
<td>17.4 (45)</td>
<td>229.137 (34)</td>
<td>89.55 (46)</td>
<td>9.846 (25)</td>
<td>1.655 (12)</td>
</tr>
<tr>
<td>Tokyo_12</td>
<td>Rumiko</td>
<td>-0.814 (15)</td>
<td>24.562 (25)</td>
<td>255.367 (18)</td>
<td>122.53 (35)</td>
<td>8.846 (40)</td>
<td>1.182 (45)</td>
</tr>
<tr>
<td>Tokyo_13</td>
<td>Yoshiko</td>
<td>-3.465 (30)</td>
<td>26.567 (13)</td>
<td>275.34 (6)</td>
<td>166.28 (15)</td>
<td>10.108 (20)</td>
<td>2.058 (3)</td>
</tr>
<tr>
<td>Tokyo_14</td>
<td>Kasumi</td>
<td>-4.412 (32)</td>
<td>30.192 (1)</td>
<td>263.436 (12)</td>
<td>165.73 (16)</td>
<td>7.993 (45)</td>
<td>1.444 (30)</td>
</tr>
<tr>
<td>Tokyo_15</td>
<td>Yuka</td>
<td>-4.678 (34)</td>
<td>22.995 (31)</td>
<td>231.109 (32)</td>
<td>131.90 (28)</td>
<td>10.526 (10)</td>
<td>1.499 (28)</td>
</tr>
<tr>
<td>Kyoto_1</td>
<td>Akari</td>
<td>-13.27 (46)</td>
<td>22.762 (33)</td>
<td>299.376 (1)</td>
<td>216.47 (2)</td>
<td>11.495 (3)</td>
<td>2.318 (2)</td>
</tr>
<tr>
<td>Kyoto_2</td>
<td>Nao</td>
<td>-1.832 (21)</td>
<td>23.197 (30)</td>
<td>255.85 (16)</td>
<td>262.49 (1)</td>
<td>10.763 (8)</td>
<td>1.613 (15)</td>
</tr>
<tr>
<td>Kyoto_3</td>
<td>Chie</td>
<td>-8.445 (44)</td>
<td>26.828 (9)</td>
<td>244.44 (23)</td>
<td>147.14 (21)</td>
<td>9.65 (26)</td>
<td>1.697 (11)</td>
</tr>
<tr>
<td>Kyoto_4</td>
<td>Futaba</td>
<td>2.893 (5)</td>
<td>18.358 (42)</td>
<td>225.253 (38)</td>
<td>108.16 (40)</td>
<td>172.49 (10)</td>
<td>1.837 (6)</td>
</tr>
<tr>
<td>Kyoto_6</td>
<td>Mika</td>
<td>-6.013 (39)</td>
<td>28.72 (3)</td>
<td>264.665 (11)</td>
<td>140.12 (25)</td>
<td>11.231 (4)</td>
<td>1.504 (25)</td>
</tr>
<tr>
<td>Kyoto_7</td>
<td>Sayumi</td>
<td>-0.197 (13)</td>
<td>28.455 (5)</td>
<td>284.95 (5)</td>
<td>170.09 (12)</td>
<td>10.492 (11)</td>
<td>1.305 (36)</td>
</tr>
<tr>
<td>Kyoto_8</td>
<td>Mutsuki</td>
<td>-6.393 (40)</td>
<td>21.677 (36)</td>
<td>243.35 (24)</td>
<td>173.85 (9)</td>
<td>11.958 (33)</td>
<td>1.594 (17)</td>
</tr>
<tr>
<td>Kyoto_9</td>
<td>Sakura</td>
<td>-3.136 (26)</td>
<td>28.437 (6)</td>
<td>258.83 (14)</td>
<td>145.15 (22)</td>
<td>8.545 (42)</td>
<td>1.166 (46)</td>
</tr>
<tr>
<td>Kyoto_10</td>
<td>Eri</td>
<td>2.954 (4)</td>
<td>28.719 (4)</td>
<td>249.682 (50)</td>
<td>120.01 (37)</td>
<td>11.958 (33)</td>
<td>1.594 (17)</td>
</tr>
<tr>
<td>Kyoto_11</td>
<td>Rika</td>
<td>-2.591 (24)</td>
<td>28.277 (7)</td>
<td>213.861 (41)</td>
<td>129.70 (29)</td>
<td>10.133 (19)</td>
<td>1.502 (26)</td>
</tr>
<tr>
<td>Kyoto_12</td>
<td>Orie</td>
<td>-4.641 (33)</td>
<td>20.495 (38)</td>
<td>260.878 (13)</td>
<td>141.50 (24)</td>
<td>10.406 (21)</td>
<td>1.309 (35)</td>
</tr>
<tr>
<td>Kyoto_13</td>
<td>Tomomi</td>
<td>1.291 (10)</td>
<td>26.042 (14)</td>
<td>260.878 (13)</td>
<td>141.50 (24)</td>
<td>11.91 (1)</td>
<td>1.625 (14)</td>
</tr>
<tr>
<td>Kyoto_14</td>
<td>Sae</td>
<td>-1.318 (18)</td>
<td>26.82 (10)</td>
<td>218.531 (40)</td>
<td>143.94 (23)</td>
<td>8.936 (17)</td>
<td>1.713 (8)</td>
</tr>
<tr>
<td>Kyoto_15</td>
<td>Yumi</td>
<td>-0.692 (14)</td>
<td>23.729 (28)</td>
<td>255.547 (17)</td>
<td>160.55 (17)</td>
<td>8.936 (17)</td>
<td>1.713 (8)</td>
</tr>
</tbody>
</table>

*Parentheses indicate ranks of the values among the 46 participants.
**The top 10 rankings and the bottom 10 rankings are bolded.

### 6.6. Intra-Regional Variation in Breathy Voice

#### 6.6.1. Meaning of Breathy Voice

216
In Chapter 3, although the Osaka participants were found to generally use more breathy voice than the Kyoto participants, I argued that it is more important to examine how breathy voice co-occurs with other features is used rather than the overall degrees of breathiness since breathy voice can be caused by various factors, including physiological ones. In fact, the ways breathy voice is used was significantly different between Kyoto and the other two regions. In Kyoto, breathy voice was robustly correlated with low intensity and high Fo, while the same correlation was not found in Tokyo and Osaka. In investigation of the meaning indexed by H1-H2, this section reemphasizes the point that breathy voice gains meaning only in connection with co-occurring linguistic features. In what follows, I will examine the stylistic practices of two participants from each region who exhibit a contrast from each other in H1-H2. As is outlined in Chapter 3, the most widely accepted meaning of breathy voice is femininity, but more specific meanings have been proposed by sociophonetic studies, such as a mature femininity (Starr 2010), a pretty and petite female figure (Addington 1968), an upper-class femininity (Pittam 1987), sexual arousal (Hall 1995, Kajino and Moon 2011, Guzman et al. 2013), sadness and tenderness (Guzman et al. 2013), and politeness (Ito 2005). As will be illustrated below, the particular meaning of breathy voice drawn from my data is closely aligned with the meaning proposed by Starr (2010), “mature femininity.” In this section, I will analyze six speakers (two from each region) of participants that show contrasts in their H1-H2 values and demonstrate how the participants exhibiting higher H1-H2 values commonly present more mature feminine figures. The two speakers from Tokyo are actually conversational partners, and the rest of the speakers analyzed from Kyoto and Osaka were not paired as conversational partners. In the latter case, I will only be looking at one speaker from each of the four paired-conversational recordings.

*Jun and Mei from Tokyo*
Jun and Mei are both juniors at Tokyo Rose University and also both third generation Korean Japanese who were born and raised in Tokyo for their entire lives. Their first language is Japanese, and they had not learned Korean until they took a Korean language course at college. Jun and Mei have been very close friends since high school, but their linguistic and non-linguistic styles are strikingly different. As is seen in Table 6.2, while the mean COG of /s/ are similar between Jun and Mei, Jun’s mean H1-H2 value is much higher than Mei’s, which indicates that Jun uses more breathy voice than Mei. Jun’s breathy voice is situated in her speech along with low F0 and narrow F0 range as well as a low degree of final lengthening, whereas Mei’s non-breathy voice is produced with high F0, wide F0 range, slow speech rate, and a high degree of final lengthening.

These contrasting features in Jun and Mei’s speech reflect and help construct their contrasting types of femininities. When I first met Jun and Mei together, the contrast in their fashion styles struck me immediately. While Jun likes a local Tokyo-based clothing brand, which is marketed as “edgy,” “sexy,” and “glamorous,” (see Figure 6.2) the kind of fashion style Mei chooses is commonly called *mori gaaru* (‘forest girl’) style, which uses soft and natural fabrics with a light and neutral color scheme, which is characterized by the key words, “natural,” “feminine,” and “girly” (see Figure 6.3).
In the interview with me, Jun explained that her identity was rooted in memories of her adolescence in Excerpt 6.1\(^{39}\).

Excerpt 6.1 (In the response to the question “Has anyone said anything positive or negative about the way you speak?”)

1. Jun: Well... Like I sound too manly. And... yea, too manly.
2. Int: I wonder what makes [your speech] sound manly..
3. Jun: I used to say ore
4. Int: Mm hmm, why were you using ore?
5. Jun: I didn’t like being a girl...
6. Int: I see.
7. Jun: For awhile, I was like “Why am I a girl?”

\(^{39}\) Original Japanese transcriptions of all the excerpts presented in this chapter are attached as Appendix III at the end of this dissertation (p. 270).
9. Jun: Junior high. Boys looked really cool to me.
10. Int: Mm Hmm, yea.
11. Jun: Then, I had a crush, and it changed, but until then, boys looked simply cool. And boys are like.. somehow, devoting 100% of their energy to playing sports, studying, or doing anything, but girls worry a lot about how they are seen, you know? I hated it, so I was like “Oh my god, boys are so cool, why am I not a boy?” Also since I was taller than other girls, I was always described by other girls as “cool/handsome” rather than “cute” or “pretty.”

Since Jun was a child, she had been described as “kakkoii ‘cool’ or ‘handsome’” rather than “kawaii ‘cute’” because of, she believes, her tall height (Lines 15-17). But these impressions of Jun were also backed up by her active rejection of typical girls’ practices and nontraditional linguistic usages such as referring to herself with the masculine form ore (Line 3). Today, Jun struggles with the tendency to be the one “wearing the pants” in relationships while expressing a serious concern for her mentally unstable ex-boyfriend who still seeks her emotional support.

In contrast with Jun, Mei has been described as “osanai ‘immature’” and “kodomoppoi ‘childish’.” In Excerpt 6.2, Mei tells me that she was advised by older people to change her manner of speaking (Lines 2-4). Though Mei initially says that she does not know what makes her speech sound immature, after Jun points out Mei’s slow speech rate (Line 7), Mei shows awareness of her own slow speech rate by saying “I can’t speak fast” (Line 11). In Lines 14-16, Jun emphasizes the contrast between herself and Mei in their non-linguistic practices such as appearance.

220
Excerpt 6.2 (In response to the question “Has anyone said anything positive or negative about the way you speak?”)

1. Mei: Yes, I have, but like.. well, for me, it’s not... really I don’t know why,
2. but well: .. before, when I met people above me or older people:, It was
3. said:, “the way you speak sounds too immature relative to your age or
4. you speak in a childish manner, so it’s better to change the way
5. you speak” <laugh>
6. Int: Hm, I wonder why?
7. Jun: Is-Isn’t it because you speak slowly? Because you drawl.
8. Mei: <laugh> Ah I wonder if that’s the case. But I totally have no idea: like
9. often.. people say that to me...
10. Jun: Because [your speech] lacks dynamism [in the rhythm]?
11. Mei: No dynamism <laugh>. But I can’t fix it:. I can’t speak fast:.
12. Jun: But I think the impression has something to do with her looks.
13. Int: How?
14. Jun: Since Mei talks like that with her looks, it’s considered immature but if [I]
15. talk like a kid with these looks <pointing to her own face>, then people
16. would describe it as burikko.

At the same time, in Lines 14-16, Jun makes an interesting point about style: If Jun spoke like Mai, it could be interpreted as a bogus gender performance (burikko) rather than as more “authentic” age-related identity work (child-likeness or immaturity). Although age is not separable from gender since it holds an important orthogonal dimension of gender identities (e.g. Eckert 1996, Kajino and Podesva 2007, Stuart-Smith 1999), Jun’s point highlights that
linguistic and non-linguistic aspects of style should work in concert to form an “authentic” style since people make different assumptions about whether or not the use of a linguistic feature is “authentic” based on the alignment with other co-occurring linguistic features as well as nonlinguistic features of the speaker’s style. This effect of people’s assumptions in interpreting meanings of linguistic variants is also reported in Podesva, Jamsu, and Callier (forthcoming). In their perception study on politicians’ speech, Podesva et al. found that listener interpretations of meanings of released /t/ vary according to the public images or identities of the politicians/speakers to which listeners are pre-exposed.

The Tokyo participants examined in this section, Jun and Mei, construct distinctive gender identities linguistically and non-linguistically. Jun presents herself as a strong-willed and mature woman through the clothes she wears, stories she tells, and linguistic features she uses which include a high degree of breathiness along with low F0, a narrow F0 range, and a low degree of final lengthening. Mei constructs a very different feminine identity from Jun’s which is described as “childish” and “immature,” using non-linguistic resources such as girly clothes and linguistic resources including a low degree of breathy voice, high F0, a wide F0 range, low speech rate, and a high degree of final lengthening.

Miya and Nanako from Osaka

<table>
<thead>
<tr>
<th>Region_no.</th>
<th>Speaker</th>
<th>H1-H2 (dB)</th>
<th>/s/ COG (erb)</th>
<th>F0 (Hz)</th>
<th>F0 Range (Hz)</th>
<th>Speech Rate (mora/sec)</th>
<th>Final Lengthening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osaka_9</td>
<td>Miya</td>
<td>1.653 (9)</td>
<td>25.7 (16)</td>
<td>218.698 (39)</td>
<td>97.90 (44)</td>
<td>9.901 (24)</td>
<td>1.42 (32)</td>
</tr>
<tr>
<td>Osaka_11</td>
<td>Nanako</td>
<td>-6.494 (41)</td>
<td>25.884 (15)</td>
<td>267.485 (10)</td>
<td>213.90 (3)</td>
<td>9.338 (32)</td>
<td>1.358 (34)</td>
</tr>
</tbody>
</table>

Table 6.3. Summary of inter-speaker variation between Miya and Nanako from Osaka
Miya and Nanako from Osaka show a contrast similar to that observed between Jun and Mei. Among the four speakers, Miya patterns similarly to Jun, while Nanako shows a pattern like Mei’s. It is also clear that the mean COG values of /s/ present a regional distinction among the participants, with lower mean values indicating the Tokyo region and higher mean values indicating the Osaka region. Just like Jun and Mei’s speech styles do, Miya and Nanako’s speech styles differ from each other in H1-H2 as well as F0, and F0 range (see Table 6.3). Again, these linguistic features work with their non-linguistic styles and build the contrast between the mature style of Miya and the immature style of Nanako.

Miya is a fourth-year student and a member of Osaka Lily University’s cheerleading team. Their cheerleading team is a frequent winner of college national championships, and constitutes one of the most visible and strongly tied communities of practice on the campus. Because of the strong bonds and visibility on campus, other students find members of the cheerleading team “scary (kowai)” and “hard to approach (chikayori nikui).” Members of the cheerleading team always wear their training clothes with the team logo and carry the same bags printed with the team name. The members congregate outside of training, for example, in the school cafeteria and the school buses, but the hierarchy between freshmen, sophomores, juniors, and seniors is strictly maintained. Within the team, Miya is one of the core members because of her outstanding cheerleading skills, and she even went to a world competition with the all-Japan team. She has many younger “admirers,” and during my fieldwork, a few times, I encountered scenes where Miya was advising younger teammates as an older experienced colleague, and the younger teammates were consistently using polite linguistic forms toward her. In the following excerpt from the conversation recording, Miya tells her teammate Kana, who is also a senior, how hard a freshman member, Atsuko, is trying on the team. Like the rest of their conversations, in this excerpt, Kana tends to serve as a good listener and agree with Miya’s statements.
Excerpt 6.3 (Miya told Kana that Atsuko is working five days a week at restaurants to pay her own cheerleading expenses)

1. Kana: 5 days a week!
2. Miya: Yea, she ‘aitsu’ is so crazy.
3. Kana: Wow, how does [she] do that... Plus, [she] is still in her first year and has many classes.
4. Miya: Yea, there are a lot of morning classes.
5. Kana: I can’t believe that [she] does that. I would totally fail at it.
7. Kana: Aren’t all the freshmen this year tough?
9. Kana: [They] are doing okay in training, so I guess that’s no problem.
10. Miya: They ‘anoko-ra’ are not getting to the point where [they] have to struggle yet.
12. Miya: [They] can do anything now.
13. Kana: Yea, and [they] are already doing well.
14. Miya: I miss those days.

This interaction between Miya and Kana shows how Miya positions herself above Atsuko and other freshman team members in terms of maturity and the hierarchical status within the team. While Kana avoids using any referential terms40, Miya refers to Atsuko and other freshman members as aitsu ‘that girl’ (Line 2) and anoko-ra ‘those girls/kids’ (Line 11), which both carry a

40 Note that Japanese is a pro-drop language in which pronouns are often omitted.
connotation of superior authority as well as familiarity toward the speaker (Lee and Barbeau 2012). Particularly, *aitsu* indexes toughness by displaying her power and authority, while the equivalent form *anoko* ‘that girl/kid’ (singular form of *anoko-ra*) connotes a tender and motherly image (Lee and Barbeau 2012). Miya’s utterances in Lines 11 and 16 further enhance her positioning as a mature and advanced team member.

On the other hand, Nanako is a first-year student in the recently founded nursing department of Osaka Lily University. Nanako does not belong to a sports or cheerleading team, but she enjoys going to watch collegiate American football games where her boyfriend plays. Her email address contains her boyfriend’s initials and uniform number, and in the conversation with her friend Ana, who is also a first-year nursing student, she openly talks about her relationship. In Excerpt 6.4, Nanako tells Ana about her post-graduation plan.

Excerpt 6.4 (Nanako and Ana are looking at magazines on careers)

1. Nana: I *'Nana'* really think of getting married without getting a job.
2. Ana: Yea, sounds good to me *'atashi'* too.
3. Nana: Seriously, [I] will get married and not work.
4. Ana: But it’s so rare to find someone like that in our age, you know?
5. Nana: Yea, so rare. Let’s say, our boyfriends find jobs, and we *'Nana-ra'*
6. are married without getting jobs. And then, the places [we] live
7. wouldn’t be [our] own home, but [our] parents’ home, until enough
8. money is saved.
9. Ana: If the husbands are our age, that’s the case.
11. Ana: Right.
12. Nana: [I] don’t want to ask for parents’ help.
13. Ana: But if were like that, [I] don’t think I would become a stay-home-
14. wife for sure, not if the situation were like that.
15. Nana: But even if [we] go work, [we] wouldn’t make that much money...
16. Ana: Yea, maybe it would just be a part-time job.

Throughout the conversation recording as well as the interview with me, Nanako exclusively uses Non-Pronominal Self-Reference (NPSR), which is an unconventional form of self-reference, in which the speaker refers to oneself with her first name (Kajino and Podesva 2007, Kajino 2011). NPSR is a typical feature of Japanese “baby talk” and commonly used by young females to connote childishness and femininity. As is shown in Lines 1 and 5 of Excerpt 6.4, Nanako refers herself with her first name Nana, while Ana uses the conventional feminine referential form atashi (Line 2). In the excerpt, Nanako’s dream of becoming a stay-at-home-wife is in conflict with the reality of modern times. Ana proposes a more realistic plan in Lines 13-14, and Nanako struggles with the compromise.

**Eri and Mika from Kyoto**

<table>
<thead>
<tr>
<th>Region_no.</th>
<th>Speaker</th>
<th>H1-H2 (dB)</th>
<th>/s/ COG (erb)</th>
<th>F0 (Hz)</th>
<th>F0 Range (Hz)</th>
<th>Speech Rate (mora/sec)</th>
<th>Final Lengthening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyoto_10</td>
<td>Eri</td>
<td>2.954 (4)</td>
<td>28.719 (4)</td>
<td>249.685 (20)</td>
<td>120.01 (37)</td>
<td>9.158 (33)</td>
<td>1.594 (17)</td>
</tr>
<tr>
<td>Kyoto_6</td>
<td>Mika</td>
<td>-6.013 (39)</td>
<td>28.72 (3)</td>
<td>239.63 (26)</td>
<td>172.49 (10)</td>
<td>10 (23)</td>
<td>1.7 (10)</td>
</tr>
</tbody>
</table>

Table 6.4. Summary of inter-speaker variation between Eri and Mika from Kyoto
Eri and Mika’s speech also contrasts in H1-H2 with Eri using more breathy voice than Mika as presented in Table 6.4. Another prominent contrast is that Eri uses breathy voice with narrow F0 range like Jun and Miya, and Mika uses non-breathy voice with wide F0 range like Mei and Nanako. As is discussed in Chapter 5, Mika is one of the five Kyoto participants who use the deaccented form of negative polar questions, but Eri never uses the feature.

Eri and Mika are both freshman college athletes. Eri is a volleyball player, and Mika is a long-distance runner on the college track & field team, but Mika’s involvement in the team is much higher than Eri’s, since the volleyball team is only active seasonally, while track & field requires daily training year round. In addition to her involvement in volleyball, Eri works at a local French cake shop as an apprentice. She is very serious about becoming a pastry chef, and her motivation for pursuing this career is, as she explained, that she feels pleasure serving others. Her serious and caring nature can be observed from her interaction with Mika in the recorded conversation. In Excerpt 6.5, Mika, who just confessed that she has never dated anyone before, asks Eri a question about relationships. The advice Eri gives to Mika, based on her own experience, (Lines10-11, 13-14) is quite insightful.

Excerpt 6.5

1. Mika: I’ve been wondering, I’ve been wondering for awhile, what do people
2. do once they are in relationship?.. Like, I’ve been thinking.
3. Eri: What an interesting person you are...
4. Mika: <laugh> Eh, because I was just wondering. Well, in romance, before
5. you get serious.. it’s the best time. Somehow like ahhh..
6. Eri: Uh... when you have a crush, right?
7. Mika: Right, I think a crush is fun, like quite fun.
8. Eri: It’s exciting.

9. Mika: Once you are in a relationship, you may fight with your boyfriend.

10. Eri: Yea, fights. But if you can’t fight, the relationship will eventually end. I think fighting isn’t bad.

11. Mika: Oh yea?

12. Eri: If you are very passive like you can’t fight, and you keep everything to yourself, then maybe you shouldn’t be in the relationship. That’s how I broke up <laugh>.

13. Mika: Oh really?


In contrast with Eri, Mika’s immaturity is shown in her view of relationships expressed in Excerpt 6.5 (Lines 1-2, 4-5). As is discussed in Chapter 5, Mika is a big fan of girls’ romantic mangas (‘comics’) which target teenaged readers, and we can see here again how Mika embraces the heterosexual ideology presented by this particular medium. The main theme of typical girls’ romantic mangas is having a crush and struggling to find love, just like Mika’s understanding of romance.

In this section, I have highlighted the stylistic practices of three participants whose speech exhibits a high degree of breathiness and three other participants whose speech shows a low degree of breathiness and argued that the former group of participants commonly presented more mature figures than the latter group. However, my point is not that a high degree of breathy voice alone indexes maturity per se, but that it contributes to the mature style when used in co-occurrence with other linguistic and non-linguistic features. A low degree of breathy voice by itself does not index immaturity either, but along with other relevant features such as a wide F0 range, it co-constructs the immature figure.
In what follows, I will highlight the effects of co-occurring features in the meaning of breathy voice. I will further break down the mature feminine figures constructed with breathy voice into two groups with more specific associated images: the soft and sensitive feminine image and the strong and bold feminine image. The soft and sensitive feminine style is co-constructed with breathy voice and higher pitch, while the strong and bold feminine style is co-constructed with breathy voice and lower pitch.

### 6.6.2. Co-Construction of Meaning through Breathy Voice and Pitch

<table>
<thead>
<tr>
<th>Region_no.</th>
<th>Speaker</th>
<th>H1-H2 (dB)</th>
<th>/s/ COG (erb)</th>
<th>F0 (Hz)</th>
<th>F0 Range (Hz)</th>
<th>Speech Rate (mora/sec)</th>
<th>Final Lengthening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyoto_10</td>
<td>Eri</td>
<td>2.954 (4)</td>
<td>28.719 (4)</td>
<td>249.685 (20)</td>
<td>120.01 (37)</td>
<td>9.158 (33)</td>
<td>1.594 (17)</td>
</tr>
<tr>
<td>Kyoto_13</td>
<td>Tomomi</td>
<td>1.291 (10)</td>
<td>26.042 (14)</td>
<td>260.878 (13)</td>
<td>141.50 (24)</td>
<td>10.104 (21)</td>
<td>1.309 (35)</td>
</tr>
<tr>
<td>Osaka_8</td>
<td>Naho</td>
<td>7.968 (1)</td>
<td>22.851 (32)</td>
<td>258.222 (15)</td>
<td>106.83 (41)</td>
<td>10.685 (8)</td>
<td>1.299 (37)</td>
</tr>
<tr>
<td>Tokyo_8</td>
<td>Jun</td>
<td>2.591 (7)</td>
<td>20.991 (37)</td>
<td>199.26 (44)</td>
<td>122.84 (34)</td>
<td>9.641 (27)</td>
<td>1.22 (41)</td>
</tr>
<tr>
<td>Osaka_9</td>
<td>Miya</td>
<td>1.653 (9)</td>
<td>25.7 (16)</td>
<td>218.698 (39)</td>
<td>97.90 (44)</td>
<td>9.338 (32)</td>
<td>1.358 (34)</td>
</tr>
<tr>
<td>Osaka_16</td>
<td>Yuriko</td>
<td>6.775 (2)</td>
<td>22.094 (35)</td>
<td>202.327 (43)</td>
<td>103.18 (42)</td>
<td>10.365 (14)</td>
<td>1.55 (20)</td>
</tr>
</tbody>
</table>

Table 6.5. Summary of inter-speaker variation across participants with high degrees of breathy voice

Table 6.5 shows the linguistic patterns of six participants, including three of the participants discussed in the previous section (Eri, Jun, and Miya) and three additional speakers, Yuriko, Tomomi, and Naho. These six participants are all breathy speakers, but the first and second groups contrast in their mean Fo. The mean Fo values of the first three (Eri, Tomomi, and Naho) are higher, and those of the last three (Jun, Miya, and Yuriko) are lower. In this section, I argue that breathy voice is used in the construction of different types of maturity based on its placement with different Fo levels. Whereas breathy voice positioned with higher Fo helps to
create a traditional femininity by indexing sensitivity, breathy voice positioned with lower Fo constructs unconventional femininities that present strength and boldness.

The use of breathy voice and higher Fo by Eri, Tomomi, and Naho accords with some qualities they display that are in common with the traditionally feminine image of the *ryosai kembo* (good wife, wise mother). For example, Eri’s involvement in the typical “feminine” practice, baking, shows her nurturing and caring character. Her responses to Mika’s question about relationships (Excerpt 6.5) also manifest her sensitivity and maturity, which also presents a conventional feminine figure. Recall that a conventional/traditional feminine figure indicates the *ryosai kembo* figure. This figure presents a femininity and, at the same time, maturity. As discussed in Chapter 1, the stereotypical characteristics of a *ryosai kembo* are her “motherliness”, “gentleness”, “affection”, “modesty”, and “the spirit of self-sacrifice” (Kurasumi 2008).

The feminine style of Tomomi, a third-year student from Kyoto, is represented in her conservative fashion style. Just as she carefully picked out a collarless jacket with a pearl necklace, she is careful and conscious about her feminine identity. In her adolescence, she was so self-consciousness that it prevented her from eating in front of other people, even her own family, because she was simply “embarrassed by eating a lot (いっぱい食べるとつゆのが恥ずかしさかれいてん),” considering it to be outside traditional gender expectations (see Excerpt 6.6a in Appendix III on p. 270). Though she overcame the eating issues over the years, she is still self-conscious about her own feminine presentation. In the following excerpt, Tomomi confesses to her friend Maya (who is excluded from the analysis since she is not from the designated region) that she accidentally passed gas while on the phone with her boyfriend Ken. Even if Ken may not think less of her, as Maya says, for someone who is very conscious of her conventional feminine image like Tomomi, “just the idea of him noticing it” goes against her ideas of her feminine presentation.
Excerpt 6.6

1. Tomo: We were talking on the phone... and like... It just came out when I laughed.
2. Maya: I see.. You don’t have to worry about that at all.
3. Tomo: It was over the phone, so I thought he didn’t hear it... Do you think it was all right?
4. Maya: No problem, no problem. You worry too much. But-but even if Ken heard it, I feel like he would just laugh at it.
5. Tomo: No no no no, I just can’t stand the idea of him noticing it <laugh>.

The last participant who uses breathy voice and high Fo is Naho, a second-year student from Osaka. Naho also presents a sensitive and caring femininity while being conscious about her traditional presentation of femininity. During the interviews, Tomomi and Naho both said that when they were children, their mothers specifically told them not to use “dirty (kitanai)” and “masculine (otokoppoi)” language. On the day of the conversation recording, Naho showed up with her arm wrapped in gauze. Naho burned her arm while cooking with her boyfriend a day earlier, but she did not tell her boyfriend or her family about the injury so as not to worry them. Instead, Naho asked her friend Teru to change the gauze at school. Evidence for Naho’s orientation toward the traditional femininity is also found in my data. During the conversation recording, Naho criticizes nontraditional kimono styles portrayed in a women’s fashion magazine: “Look look, for a kimono, it’s definitely much cuter if you put your hair up. It’s a Japanese custom. In Japan, people traditionally put hair up since it is the cutest way. (見て見て、anna、浴衣はな、髪の毛上げた方が絶対かわいい。日本の行事やん。なんか日本でさ、上げて、それで上でセットアップすんのが一番かわいいのにさ)” Regarding the image of the Tokyo variety during the interview,
while associating the Tokyo variety with politeness, Naho commented on male Tokyo speakers:

“Since the [Tokyo] males speak more politely [than I do], I would feel bad [if I were with them] like ‘I’m sorry even though I am female [you are speaking more politely than I do]...’ (男のひとの方が丁寧な話し方されたら、こっち女なのになんかゴメンってなる)” This comment presupposes the traditional belief that women, including her, should speak more politely than men, and the presupposition makes her apologetic for not speaking “like a woman.”

Furthermore, Naho notes that her own variety, the Osaka variety, “probably sounds dirty (kitanai) to Tokyo people.” Combined with her comment on the Tokyo men, we observe a chasm that Naho experiences between what is considered to be “ideal” feminine speech and her perception about her own speech being “harsh” and “vulgar,” the stereotypical image attached to the Osaka variety. This chasm may be a motivation for Naho to use more breathy voice than any of the other participants in my data. While Tomomi is also conscious of the traditional femininity like Naho, this conflict was not found in Tomomi’s metalinguistic commentary, which is probably because the stereotype of the Kyoto variety does not conflict with the traditional femininity Tomomi strives to attain. Note that Tomomi is one of the only two Kyoto participants who responded positively to the question, “Do you think you speak the Kyoto variety?”

The above-mentioned characteristics of the three participants present them as ryosai-kembo-like traditionally feminine women, and their use of breathy voice and high pitch is in concert with the image. Next, I focus on the second group of participants who use breathy voice and lower pitch: Jun, Miya, and Yuriko.

Jun and Miya’s strength and boldness are evident from their personae. As is discussed in the previous section, Jun tends to be the dominant one in her relationships, and Miya acts tough by displaying her power and authority in the cheerleading team. Yuriko projects toughness by ridiculing her father for his physical height in her conversation with Akiko. Talking about the recently passed father’s day, Yuriko jokingly tells Akiko that she did not give her father a gift, in
spite of her original plan to buy him clothes, since “he is so short that nothing looks good on him (めっちゃちっちゃいから何も似合わへん).” On the other hand, Yuriko highlights her tall physicality in conversations multiple times by saying that she is taller than her father or that she avoids wearing heels, which receives mixed responses of admiration and consolation from Akiko. Interestingly, the particular discourse about avoiding wearing heels because of height is commonly found in the speech of Jun and Miya as well. Jun complains that she cannot wear heels for dates because heels would make her much taller than the average Japanese male. Additionally, after looking at a magazine article showing a recent trend of Japanese men being attracted to taller girls, Miya says, “This is a total lie. I know guys complain about it when girls are really tall (こんなん絶対嘘や、でかかったらでっかかったでグチグチ文言うくせに).” These complaints reveal that these three participants see themselves as not conforming to the hegemonic ideological construct of femininity that values a petite and dainty physique, and to them, a presentation of strength and boldness is the outcome of the negotiation with the traditional ideology as well as being an emblem of the rejection of that ideology. Finally, their use of breathy voice and low F0 present the negotiation between how they see themselves and how they believe they are expected to be seen as females.41

At the end of this section, it should be noted that there is no Kyoto participant who uses high degrees of breathiness along with low pitch. Breathy voice is used by speakers whose overall pitch is relatively high in Kyoto, whereas breathy voice can be used by speakers whose mean pitch is low in Tokyo and Osaka, which suggests that the particular kind of femininities that present maturity with strength and boldness are peculiar to Tokyo and Osaka. This point corresponds with the finding in Chapter 3 where a robust positive correlation between H1-H2

41 The meaning of breathy voice with lower Fo I am proposing in this section may somewhat correspond with the authority meaning presented in Sicoli (2010, forthcoming) though the indexical mechanism is probably different. In Lachixio Zapotec, the language Sicoli studied, lower pitch marks greater authority, and breathy voice helps to enhance the level of authority by increasing low-pitch specifications beyond the laryngeal limits of their modal pitch range.
and Fo was found in Kyoto but not in Tokyo and Osaka, which further supports the link between the ways breathy voice is used in Kyoto and the classy feminine image of the Kyoto variety.

In this section, I have explored the meaning of breathy voice by comparing styles of participants who use a high degree of breathy voice and other participants who use a low degree of breathy voice. While a ‘feminine’ and ‘mature’ meaning has been proposed, breathy voice alone does not index the meaning; rather, the combination of breathy voice and co-occurring Fo levels construct particular types of mature femininities. Now we shift our attention to the meanings of the other feature focused on in this chapter, COG of the fricative /s/.

### 6.7. Intra-Regional Variation and Meaning of Frequency Energy of /s/

Chapter 4 of this dissertation found that the COG of /s/ exhibits a significant difference between Tokyo and Osaka, with low COG representing Tokyo and high COG representing Osaka. As Table 6.1 shows, lower ranks of mean COG values are observed more among the Tokyo participants than among the Osaka participants. In Chapter 4, borrowing psychoacoustic parameters of sound quality, I proposed the perceived sound quality of sharpness for a high COG of /s/ and argued that the quality corresponds with the “harsh” image of the Osaka variety. While Tokyo and Osaka were found to be significantly different, Kyoto was not found to be statistically different in its COG of /s/ from either Tokyo or Osaka. Kyoto participants exhibit diverse patterns, displaying both high and low mean COG (see Table 6.1). In this section, I demonstrate the association between high a COG of /s/ and the meaning of sharpness through examinations of Kyoto participants’ stylistic practices. Tables 6.6 and 6.7 below present inter-speaker variation between two Kyoto participants whose speech contrasts in the mean COG of /s/.
### Table 6.6. Summary of inter-speaker variation between Sae and Mami from Kyoto

<table>
<thead>
<tr>
<th>Region_no.</th>
<th>Speaker</th>
<th>H1-H2 (dB)</th>
<th>/s/ COG (erb)</th>
<th>F0 (Hz)</th>
<th>F0 Range (Hz)</th>
<th>Speech Rate (mora/sec)</th>
<th>Final Lengthening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyoto_14</td>
<td>Sae</td>
<td>-1.318 (18)</td>
<td>26.82 (10)</td>
<td>218.531 (40)</td>
<td>143.94 (23)</td>
<td>11.91 (1)</td>
<td>1.625 (14)</td>
</tr>
<tr>
<td>Kyoto_5</td>
<td>Mami</td>
<td>-3.519 (31)</td>
<td>16.285 (46)</td>
<td>225.253 (38)</td>
<td>108.16 (40)</td>
<td>9.65 (26)</td>
<td>1.519 (24)</td>
</tr>
</tbody>
</table>

Sae and Mami from Kyoto

Sae and her friend Futaba are both fifth-year seniors because they have repeatedly failed to acquire credits by skipping too many classes. Instead of attending classes, they hang out together. Sae has planned to marry her boyfriend, who is a police officer, after graduation, but he is more concerned with her failures at school than the delay of their marriage. Sae joked to me that she has had a few small arguments over the graduation issue with her boyfriend. During the recorded conversation with Futaba, Sae consistently speaks at a very fast speech rate and in a comedic manner. In the beginning of the recorded conversation, she makes fun of herself for having very dry skin from commuting to school by scooter. After that, she conveys a series of what she calls “funny” stories from the coffee shop where she works and from some classes she is taking. Some serious issues can even be “funny” when she tells them. For example, Sae talks about a news documentary she recently watched on TV that showed foreign criminals who were secretly filming women on the beach. Sae describes the criminals as looking “very silly” and “very funny”, and the scene of the arrests was comedic to her. Her general comedic and fun-seeking attitude that seems lack seriousness, along with her language use, is criticized by her mother. In the interview, Sae told me that her mother said to her that her speech sometimes “pierces (tsukisasaru)” right into the heart, and just one word can hurt others. This criticism coincides with the perceived sharpness of the /s/ sound.
In contrast to Sae’s speech style, Mami’s friends describe her speech style as “careful” and “polite” (teineina). Mami, a third-year student, belongs to the international volunteer club on campus, but the club is not very active, so she usually spends time working at the produce section of a local grocery store after classes to pay for her tuition. She enjoys gossiping with her co-workers, most of whom are middle-aged women, during breaks because she finds it easier to communicate with older people. She particularly does not feel comfortable talking with young males, which makes her more attracted to older, middle-aged men. In the recorded conversation, her friend Nao gives Mami advice on how to meet older males and makes a plan to take her to old-fashioned bars. Mami’s shy and modest persona is also clear from the shopping anecdote she conveys to Nao during the conversation in Excerpt 6.7.

Excerpt 6.7 (Mami tells Nao about when she went shopping with her friend Ayako)

1. Mami: So Ayako brought me a shirt like “what about this?” which I thought wouldn’t look good on me at all because it looked so flashy.
2. Nao: Unexpectedly- unexpectedly, you might see something.
3. Mami: Yea, and I said like “Eh? Is this all right?” Ayako was like “yea, this is definitely good,” so I tried it on.. Then, eventually, the store clerk flattered me, like they give you lots of compliments, you know? Like “this looks good on you, great on you” and [Ayako and the staff] both looked at me smiling, so I couldn’t say no... <laugh>

In Excerpt 6.7, Mami expresses regret over her indecisive and hesitant nature. As she says in Lines 1-2 in the excerpt, she does not think that any “flashy” clothes would look good on her so she prefers a plain look, which also shows her modesty. Her linguistic style with the low
frequency energy of /s/, along with low F0 and narrow F0 range, reflects and helps display her shy and modest persona.

**Sayumi and Orie from Kyoto**

<table>
<thead>
<tr>
<th>Region_no.</th>
<th>Speaker</th>
<th>H1-H2 (dB)</th>
<th>/s/ COG (erb)</th>
<th>F0 (Hz)</th>
<th>F0 Range (Hz)</th>
<th>Speech Rate (mora/sec)</th>
<th>Final Lengthening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyoto_7</td>
<td>Sayumi</td>
<td>-0.197 (13)</td>
<td>28.453 (5)</td>
<td>264.663 (11)</td>
<td>124.71 (32)</td>
<td>8.854 (39)</td>
<td>1.575 (19)</td>
</tr>
<tr>
<td>Kyoto_12</td>
<td>Orie</td>
<td>-4.641 (33)</td>
<td>20.495 (38)</td>
<td>258.83 (14)</td>
<td>145.15 (22)</td>
<td>8.545 (42)</td>
<td>1.166 (46)</td>
</tr>
</tbody>
</table>

Table 6.7. Summary of inter-speaker variation between Sayumi and Orie from Kyoto

Sayumi and Orie’s speech are also contrastive in the COG of /s/. Sayumi is a first-year student of Kyoto Cherry University and a big fan of a boy-band. She talks about TV celebrities and TV shows with her friend Rika through most of the recorded conversation. Recall that we have examined Sayumi’s use of the deaccented negative polar questions in Chapter 5. Sayumi positions herself as an expert on the subject of boy-band idols in the conversation. For example, she evaluates Rika’s stance, in which she likes whichever idols are in her favorite TV shows as, “I think [your way] is better, since once you know too much about them [like me], you can’t help comparing or noticing various things and becoming like ‘Oh, this guy is no good!...’ (それは一番良いとおもう、色々知っちゃうと、結構みんなと比べちゃうから、いや、なんか色々と見ちゃうから、変にな、「あ、こいつあかんって！」みたいにな、感じになるから・・・).” Because of Sayumi’s expert positioning, she becomes competitive with Rika every time they disagree. Particularly, when her favorite idol is criticized, she directly confronts Rika (see Excerpt 5.7 in Chapter 5). During my fieldwork, a few other students who often hang out with Sayumi told me that they would not feel comfortable being recorded with Sayumi since they think that maintaining good conversations with her alone can be difficult. Sayumi’s expert-like, judgmental, and competitive stance is considered a
stereotypical trait of boy-band nerds, called “Jani-Ota” in Japan. The high COG of /s/ in her speech style represents and helps construct Sayumi’s competitive and slightly aggressive stance with its sharp sound.

Unlike Sayumi, Orie, who is a fourth-year student in Kyoto, never confronts her friend Yumi in the recording. Rather, Orie and Yumi’s conversation showcases a strong tie between them through several interactions. In a good period of her recorded conversation, Orie serves as a very sympathetic ear to Yumi, who recently came back from a three-week-long student teaching experience. Yumi even says to Orie, “I was thinking of you all the time during my student teaching (教育実習の間中、ずっとオリエちゃんの事考えてたよ),” which makes Orie giggle. In some parts of the story from her student teaching, Yumi becomes very emotional, thinking about how much support she received from her advisers, and Orie sincerely consoles her. They even share their common past experience of being bullied and show their support for each other.

In this section, I have investigated stylistic practices of Kyoto participants who exhibit higher COG values and those who exhibit lower COG values and enhanced the point that higher COG connotes sharpness, which was argued in Chapter 4. Close examinations have further shown that the sharpness can be interpreted as other qualities like thoughtlessness, competitiveness, and aggressiveness. Now that the meanings of breathy voice and the frequency energy of /s/ are established, I examine a pair of two participants from Kyoto who present contrasts both in breathy voice and realization of the fricative /s/.

6.8. Contrast in Both Breathy Voice and Frequency Energy of /s/

Chie and Futaba from Kyoto
Chie and Futaba show reverse patterns in H1-H2 and COG of /s/. Whereas Chie’s speech exhibits a low mean value of H1-H2 and a high mean COG of /s/, Futaba’s speech exhibits a high mean H1-H2 value and a low mean COG of /s/. For the other linguistic features, their speech is more compatible, except in the F0 range. The F0 range of Chie is much larger than the F0 range of Futaba. The contrastive linguistic patterns between Chie and Futaba help them construct contrastive styles.

Chie is a third year student and is described by other students as having a “bad mouth.” Some students witnessed Chie having a loud fight with one of her classmates. One of the witnesses believes that it happened because of a miscommunication between Chie and her classmate due to Chie’s usual bad mouth. Another student was ridiculed by Chie, when she said, “Is your head still filled with pretty flowers?” implying that the student is an airhead. During the interview, Chie told me that many of the Kyoto Cherry University students “seem like idiots (ahoppoi)”. In her recording with Sakura, Chie severely trash-talks a female idol saying, “I will beat her up if that’s the case, I’ll go to [female idol]’s office and beat her up (しばらく、そんな。[female idol] の事務所行ってしばらく行く).” Also, Chie openly talks about her body and sexuality. She does not hesitate to tell Sakura stories of sexual adventures and sex play she has had with her boyfriend. She describes her own speech style as “very manly (otoko masari)”, and her aggressive and nontraditional feminine persona is constructed with a low degree of breathiness and a high frequency energy of /s/ along with a wide pitch range and fast speech rate.
In contrast with Chie, Futaba, who was recorded with Sae and is a fifth-year senior, is a soft and peaceful homebody. When I met Futaba in mid June, she had not yet started a post-graduate job search. Considering that Japanese college students commonly start applying for jobs in the fall of their junior years and receive offers by early summer in their senior years, she was very behind schedule. In the recorded conversation, Sae tells Futaba that her mother can help Futaba get a job in life insurance sales, but she turns the offer down since life insurance sales “sounds hard (taihensoo)”. Futaba currently works at a national chain ramen shop part-time as a waitress, and staying there seems like a good option to her. The primary reason why Futaba is not applying for jobs is that she has a very strong attachment to the small local area around her home. In fact, her decision to enroll in Kyoto Cherry University was simply based on its geographical closeness to her home. While she likes feminine clothing style, coziness is also an important element of fashion to her. At the time of the interview, she was wearing a tunic with a small floral print and a pair of crocs sandals. Her peaceful and soft persona, which presents a clear contrast from Chie’s, is embodied in her breathy voice and low frequency energy of /s/.

6.9. Negotiating the Local Stereotype

*Naho and Mariyo from Osaka*

<table>
<thead>
<tr>
<th>Region_no.</th>
<th>Speaker</th>
<th>H1-H2 (dB)</th>
<th>/s/ COG (erb)</th>
<th>F0 (Hz)</th>
<th>F0 Range (Hz)</th>
<th>Speech Rate (mora/sec)</th>
<th>Final Lengthening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osaka_8</td>
<td>Naho</td>
<td>7.968 (1)</td>
<td>22.851 (32)</td>
<td>258.222 (15)</td>
<td>106.83 (41)</td>
<td>10.685 (8)</td>
<td>1.299 (37)</td>
</tr>
<tr>
<td>Osaka_12</td>
<td>Mariyo</td>
<td>-4.827 (35)</td>
<td>26.626 (11)</td>
<td>228.3 (35)</td>
<td>199.32 (6)</td>
<td>9.505 (29)</td>
<td>1.701 (9)</td>
</tr>
</tbody>
</table>

Table 6.9. Summary of inter-speaker variation between Naho and Mariyo from Osaka
As seen from previous sections, Chie and Naho are two speakers who most clearly present figures that are divergent from their local stereotypes. By using a low degree of breathy voice and a high COG of /s/, Chie from Kyoto constructs a female figure that is contrastive with the classy and soft Kyoto stereotype. Naho from Osaka displays a traditional feminine figure that is divergent from the stereotypical Osaka image. While producing a high degree of breathiness, Naho avoids using a high COG of /s/ unlike many other Osaka participants. In contrast to Naho, Tomomi, who identifies herself as a speaker of the Kyoto variety, appears to align herself with the local stereotype. In this section, I will introduce another Osaka participant, Mariyo, who shows contrast with Naho in her linguistic practices and positioning in relation to the local stereotype.

Table 6.9 summarizes Naho and Mariyo’s usage of the six features. The data on Naho’s speech that was presented earlier in this chapter is repeated here. Naho and Mariyo are both second-year students from Osaka, but Mariyo’s speech style contrasts with Naho’s style in both H1-H2 and COG of /s/. Mariyo uses a lower degree of breathy voice and a higher COG of /s/, with lower F0, a wide F0 range, lower speech rate, and longer final lengthening, than Naho. What distinguishes the two speakers is not only their linguistic use, but also their stances toward the local stereotype. While Naho struggles with the mismatch between the image of the particular femininity to which she is orienting and the Osaka local stereotype, Mariyo embraces the stereotype. In the interview, Mariyo’s strong identity as a speaker of the Osaka variety was expressed in the response to the question, “Do you think you speak the Osaka variety?” In response, she emphasized the positive answer as “Yes, definitely (gattsuri). I use [the Osaka variety] all the time, even when texting,” and noted that she had not met anyone from Tokyo personally. Her response to another interview question also reveals that she aligns herself with the stereotypical Osaka Auntie figure. Excerpt 6.8 is a part of her narrative responding to the
question about how she likes Osaka. After mentioning how friendly people in Osaka are, she began the following excerpt. Note that Kimi is Mariyo’s conversational partner.

Excerpt 6.8 (In response to the question “Do you like Osaka?” Mariyo is explaining that Osaka people are friendly.)

1. Mari: Then, well, I can very comfortably chat with anyone regardless of their age, like middle-aged men and women in their 50s and 60s.
2. They always start talking to me! They totally think of my age wrong... I think <laugh>. In fact, it’s true that I am not in my 20s in my heart,
3. definitely <laugh>.
4. Int: Are you serious!?
5. Mari: Yea, old men in their 50s and 60s start talking to me! And when I go to stores, they ask me like “What’s the great deal today?”, so I’m like “I’m not working here!” <laugh>. Even at a drug store, I was asked like “What’s on sale today?” so I-I think to myself like “I’m not a pharmacist!,” it’s everywhere, at grocery stores, some people asked me, “Which store has the best deal?” but somehow, I can answer those questions <laugh>.
6. Kimi: It’s amazing you can answer them <laugh>.

The narrative in Excerpt 6.8 clearly constructs Mariyo as an “auntie” figure. She positions herself as older (I am not in my 20s in my heart in Line 4) and claims that other people identify her as much older (They totally think of my age wrong in Line 3). Her described behaviors –
going grocery shopping and looking at advertisements from grocery stores – support her participation in auntie-like practices as well. In addition to these non-linguistic aspects of her identity practices, Mariyo’s locally oriented linguistic style exhibiting high COG of /s/ along with a low degree of breathiness is an essential component of the Osaka Auntie figure.\footnote{While Mariyo displays an “older” figure, this does not conflict with the lack of breathiness, which I have previously argued, connotes mature femininity. An Osaka Auntie may be “old” since they are middle-aged, but “maturity” is not a trait of the figure; rather, an Osaka Auntie is comedic, cheerful, and playful, which is constructed through her wide F0 range.}

Thus far, I have focused on variation in H1-H2 and/or the COG of /s/ and different ways of negotiating the local stereotypes. Before closing this chapter, I should note the limitations to understanding linguistic styles when examining only a few linguistic features. The following section will discuss the limitations.

### 6.10. Limitations of Focusing on Only Two Features

**Asako, Reina, Taeko, and Saori from Osaka**

<table>
<thead>
<tr>
<th>Region_no.</th>
<th>Speaker</th>
<th>H1-H2 (dB)</th>
<th>/s/ COG (erb)</th>
<th>F0 (Hz)</th>
<th>F0 Range (Hz)</th>
<th>Speech Rate (mora/sec)</th>
<th>Final Lengthening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osaka_2</td>
<td>Asako</td>
<td>-2.685 (25)</td>
<td>25.169 (20)</td>
<td>269.598 (7)</td>
<td>189.20 (7)</td>
<td>8.982 (36)</td>
<td>1.953 (4)</td>
</tr>
<tr>
<td>Osaka_3</td>
<td>Reina</td>
<td>-3.438 (29)</td>
<td>25.195 (19)</td>
<td>286.474 (3)</td>
<td>156.75 (18)</td>
<td>11.779 (2)</td>
<td>1.217 (42)</td>
</tr>
<tr>
<td>Osaka_5</td>
<td>Taeko</td>
<td>-0.843 (16)</td>
<td>25.244 (17)</td>
<td>197.108 (45)</td>
<td>123.04 (33)</td>
<td>10.333 (16)</td>
<td>1.208 (43)</td>
</tr>
<tr>
<td>Osaka_6</td>
<td>Saori</td>
<td>-1.779 (20)</td>
<td>24.868 (23)</td>
<td>211.474 (42)</td>
<td>102.34 (43)</td>
<td>9.572 (28)</td>
<td>1.239 (40)</td>
</tr>
</tbody>
</table>

Table 6.10. Summary of inter-speaker variation between four Osaka participants

Compared to Tokyo and Kyoto, as is noted earlier, Osaka does not exhibit much variation in the H1-H2 and COG of /s/ means. However, a more inclusive analysis along with the mean Fo, Fo range, speech rate, and final lengthening illuminates great inter-speaker variation among the
Osaka participants. For example, as Table 6.9 shows, the mean values of H1-H2 and /s/ COG among Asako, Reina, Taeko, and Saori are very similar to one another, despite the fact that the use of other linguistic features as well as their personae diverge greatly. Their linguistic styles are better represented in other linguistic features such as Fo, Fo range, speech rate, and final lengthening, rather than in H1-H2 and COG of /s/.

Among the participants in this study, Asako spoke to me the most politely using correct forms of honorifics in face-to-face interactions as well as in text messages. Asako is a first-year student in the English department of Osaka Lily University who graduated from the high school attached to the college. She chose the school because her mother is an alumnus, and studying abroad is one of her goals during college. In the recorded conversation, Asako talks about her loving dog as well as fashion styles and make-up she wants to try now that she is in college. Reina is a first-year student in the nursing department. In contrast with Asako, who does not wear make-up, Reina usually puts on make-up perfectly and even wears artificial eyelashes and colored contact lenses. Among the Osaka participants, Reina is the only participant who expresses a desire to move to Tokyo after graduation, because of her interest in the entertainment industry. The speech of Asako and Reina exhibits similarities in the H1-H2, the COG of /s/, and the Fo level, but differences in the Fo range, speech rate, and final lengthening. The Fo range and final lengthening are greater in Asako’s speech, while Reina’s speech rate is much faster than that of Asako’s. These linguistic differences capture Asako’s calm and softer persona and Reina’s fast-paced confident persona.

Contrasting with the speech of Asako and Reina, Taeko and Saori’s speech exhibits much lower Fo levels and narrower Fo ranges, while their H1-H2 and COG of /s/ display similarities.

---

43 Asako’s slow speech rate patterns more closely to the Tokyo participants. In the interview with Asako, she noted that some of her friends have commented on her speech as “not very Osaka-like”. According to her, it is because her parents do not speak in a “strong Osaka accent” even though her parents are both from Osaka. In terms of the three target linguistic features this dissertation has examined, I did not find Asako’s speech particularly divergent from the rest of the Osaka participants, but her slower speech rate might give her friends the impression of not being “very Osaka-like.”
Taeko and Saori, who have been close friends since the 3rd grade, are third-year students in the Japanese department, and they both belong to the college calligraphy club. Taeko and Saori make fun of each other for not dressing well. They decide dates for what they called “the pajama practice days (paja-ren)”, where they go to school and calligraphy practice in relaxed and comfortable clothes like sweat pants. As is clear from Saori’s comment, “If I had extra time to pair clothes and make-up together, I would instead spend it by staying in bed longer”, they do not agree with the common gendered practices where girls invest much time and energy into becoming attractive for men. In the recorded conversation, when looking at women’s fashion magazines, they repeatedly comment on how non-functional and uncomfortable those feminine outfits look. Furthermore, based on the strong belief that men should not care much about their looks, they severely criticize men’s dyed and long hairstyles as well as flashy pointed-toe shoes portrayed in men’s fashion magazines. These commentaries from Taeko and Saori regarding young fashion cultures reveal that the “pajama practice days” are their silent protest against society’s obsession with particular looks and fashion styles. At the same time, their linguistic styles with low pitch levels and narrow pitch ranges are the audible expressions of disdain for the mainstream gendered ideology.

This section has revealed that the four Osaka participants (Asako, Reina, Taeko, and Saori) present distinctive gender identities and practices, even though their speech exhibits similarities in the H1-H2 values and the COG of /s/. This shows the limitation of understanding one’s style focusing on only two linguistic features and suggests that linguistic features that contribute to the construction of styles for some speakers are not necessarily used as stylistic features by other speakers. Therefore, more inclusive analysis of linguistic features is needed for a comprehensive understanding of the styles of all the participants in this study.
6.11. Summary

While Chapters 3-5 highlighted regional variation across Tokyo, Kyoto, and Osaka, this chapter has shed light on intra-regional variation. By examining how breathy voice and the realization of /s/ are used along with other linguistic and as non-linguistic features in individual participants’ stylistic practices, I have explored social meanings of the variants. Examination of individual styles revealed that higher degrees of breathiness helps the participants display a mature feminine figure, which is aligned with Starr’s (2010) finding. The mature feminine figure has been further divided into two feminine types based on whether the breathy voice is situated with high mean pitch or low mean pitch. Whereas breathy voice used by higher Fo speakers connotes the traditional ryosai kembo femininity indexing softness and sensitivity, breathy voice used by lower Fo speakers constructs unconventional femininities that present strength and boldness.

Whereas the realization of /s/ distinguishes the Tokyo regionality and the Osaka regionality, I have also demonstrated the roles that the association of high frequency energy of /s/ and the sound image of sharpness play in construction of femininities. Examinations of stylistic practices by Kyoto speakers have revealed that the sound image of sharpness can derive associated qualities such as thoughtlessness, competitiveness, and aggressiveness. This suggests that linguistic resources that are considered to be regional features can simultaneously connote other meanings that can be used to construct certain types of femininities, while the meanings contribute to construct gendered local stereotypes. Though the stereotypical image of the Osaka variety is the “harsh”, “vulgar”, and “auntie” figure, individual stylistic practices of Osaka participants like Naho do not correspond with the stereotypical figure. To disalign from the stereotypical Osaka image, Naho avoids using high a COG of /s/, an Osaka regional feature, that can be interpreted to carry meanings such as aggressiveness that associate with the local stereotype. Naho’s metalinguistic commentary on the Osaka variety clearly shows her struggle. Regardless of her orientation toward the conventional femininity, her regional linguistic
features prevent her from attaining the conventional feminine figure because of the stigmatized image of the regional variety. This chasm motivates Naho to use more breathy voice and lower frequency energy of /s/ than most of the participants in this study.

Finally, although the mean values of the metrics H1-H2 and COG of the fricative /s/ in this chapter have not exhibited much variation among the Osaka participants compared to the Tokyo and Kyoto participants, it is not the case that Osaka women are more homogeneous. Rather, this reveals that breathy voice and realization of /s/ cannot encapsulate individual stylistic practices, or their social meanings, in their entirety. Therefore, this chapter suggests that more comprehensive linguistic analysis is needed to fully understand the stylistic practices of the Osaka participants.
CHAPTER 7

Conclusion: Food for Thought

7.1. Overview

This dissertation is intended to bridge the gap between the two major arenas of Japanese sociolinguistics: Japanese Dialectology and Japanese Women’s Language. While both fields have attracted a great deal of scholarly attention, these fields have developed independently from each other. It is, however, worth noting that the lack of cross-field study does not mean that the concerns of the two fields never overlap. As is revealed in Chapter 1, there are contrastive gendered stereotypes attached to the language varieties of Tokyo, Kyoto, and Osaka, such as the “cold” and “distant” but “cute” Tokyo image, the “classy geisha”-like Kyoto image, and the “harsh auntie”-like Osaka image. These images are by-products of the intersection of region and gender, which are the primary concerns of the studies of Japanese Dialectology and Japanese Women’s Language.

Using these regional stereotypes as a starting point, I have attempted to uncover the mechanism whereby particular gendered images are associated with local varieties as well as how women from those regions negotiate their local stereotypes. By taking a multidimensional approach, I have examined cross-regional and intra-regional patterns of phonetic variation. In addition to the patterning of individual features, constellations of features have also been investigated. Examination of the three phonetic features that are not associated with the conventional Japanese gendered ideology have revealed how Japanese women from different regions construct femininities by deploying available linguistic resources in convergence with or divergence from the stereotypical local character types. In this conclusion chapter, I will summarize Chapters 1-6 and note contributions and limitations of this dissertation.
7.2. Chapter summary

Chapter 1 provided a review of the socio-historical background of the three regional varieties. I first surveyed the historical and current states of Japanese regional language varieties, including the governmental obsession with the Western notion of the “standard language” during the modernization period, and an overview of Japanese dialectological studies. By comparing studies in Japanese Dialectology with Western studies of “standard” varieties and women’s speech, I highlighted the point that gender has not been closely examined in Japanese Dialectology. Then, I introduced the key framework of Japanese gendered language, called Japanese Women’s Language (JWL). The complex interplay of the histories of national modernization and language standardization in the establishment of the JWL ideology was discussed in addition to previous studies of JWL. Once the background of Japanese dialectology and gendered language ideology was established, I presented the contrastive stereotypes of the three regional varieties based on the metalinguistic commentaries obtained from the participants of my study. The stereotypes portrayed by the commentaries are the boring, standard, but cute Tokyo variety, the classy and calm geisha-like Kyoto variety, and harsh and vulgar aunt-y-like Osaka variety. The latter half of the chapter focused on the theoretical and methodological approaches of the study, particularly the third wave of variation study (Eckert 2012).

Chapter 2 described the data and data collection methods. The three sets of data used in this study are 1) 25 dyadic naturalistic conversations recorded with 15 Tokyo speakers, 15 Kyoto speakers, and 16 Osaka speakers, 2) interviews with the participants, and 3) notes from the mini-ethnography. The conversations were recorded between pairs of two close friends without my presence. After the conversation recordings, interviews were conducted for each pair to
obtain the participants’ demographic information and to ask them about their views on language and gender. A mini-ethnography was conducted at each research site in order to familiarize myself with the sites, the participants, and their communities of practice. This chapter also presented descriptions of the three research sites, Tokyo Rose University, Kyoto Cherry University, and Osaka Lily University, and the 46 participants.

Chapter 3 investigated cross-regional patterns of breathy voice in the Tokyo, Kyoto, and Osaka regions. A significant point to be noted from this chapter is that a linguistic feature that is usually associated with gender like breathy voice can also display regional variation. Though the spectral analysis found that the Osaka participants generally use more breathy voice than the Kyoto participants, I argued that it is crucial to examine the co-occurring features since breathy voice gains meaning only in connection with co-occurring features. In fact, the interactions between degrees of breathiness and intensity or Fo levels exhibited significantly different patterns across the regions. First, whereas breathiness correlated strongly with low intensity in Kyoto, degrees of breathiness is rather consistent regardless of intensity in Tokyo and Osaka. Second, while the breathiness of the Kyoto speech correlated with higher Fo levels, the same correlation was not found in Tokyo and Osaka. Again, breathiness of Tokyo and Osaka speech was consistent across the varied levels of Fo. The pattern of Kyoto speech where breathy voice occurred at a lower intensity was unsurprising as breathy voice is associated with a reduction in acoustic intensity (Fischer-Jørgensen 1967, Thongkum 1988, Traill and Jackson 1988). However, what was striking is the co-occurrence of breathy voice and high Fo since breathy voice typically occurs with low Fo due to its articulatory configuration (Hombert, Ohala, and Ewan 1979: 47-48, Laver 1991: 203, Sicoli 2010). This suggested that the use of breathy voice with higher pitch in Kyoto is a consequence of social motivation and not physiologically rooted. While the Kyoto speakers produce breathy voice with linguistic features that indicate traditional femininity (i.e. low intensity and high pitch), the use of breathy voice by the Tokyo and Osaka speakers is not
associated with those features, which shows that femininities are constructed differently in Kyoto than they are in the other two regions.

While Chapter 3 found that a feature that carries gender meanings can present regional differences, Chapter 4 revealed that a feature that distinguishes regions can simultaneously connote social meanings that can be used in gender construction. In this chapter, I examined variation in realizations of the fricatives by the participants through analysis of spectral measurements (spectral peaks, COG, SD, skewness, and kurtosis) and intensity. The regression analyses found that Osaka exhibits higher COG, higher SD, lower skewness, and lower kurtosis than Tokyo, while Kyoto remained in between the two. To understand these acoustic measures more intuitively, I drew on basic notions of psychoacoustics, such as sharpness and clarity. The notion of sharpness very well described the Osaka participants’ use of variants whose energy was concentrated at higher frequencies, shown with higher COG and lower skewness. The sharp fricatives of Osaka correspond with the harsh image of the Osaka variety. On the other hand, with the lower SD and higher kurtosis, the fricatives produced by the Tokyo speakers exhibited clear and precise articulations. Also, clarity is a crucial component of the “standard” image of the Tokyo variety.

Chapter 5 highlighted the use of the deaccented form of negative polar questions (NPQs) across the three regions and explored discursive roles and social meanings of the feature in construction of femininity and locality. As previous literature has documented, this intonational pattern is enregistered as a Tokyo feature. The fact that the general public in Japan is aware of the specific regional indexicality makes this feature more salient in construction of local identities than the features investigated in Chapters 3 and 4. As is expected, the regression analysis of my data found that the Tokyo participants use the feature more than the participants from the other two regions. Based on previous literature on NPQs, I argued that in Tokyo, the deaccentuation of Japanese NPQs indicates a specific interactional stance where speakers’
epistemic and evidential biases are asserted simultaneously expressing a lack of commitment in order to show a lack of vulnerability to disagreements, which constructs “urban cuteness,” the symbolic characteristic of the “ideal” young Tokyo female figure. However, what was particularly surprising from the patterns found in my data was that five out of the 15 Kyoto participants and one Osaka participant also used the deaccented form in addition to the conventional/local accented form. Due to the regional indexicality attached to the deaccented form, I argued that the feature carries additional meaning in Kyoto and Osaka, namely the Tokyo-centric ideology, which revolves around the image of the “typical” young Tokyo female figure. In the discussion section, close examinations of unfolding conversations illuminated how those five Kyoto and one Osaka participants who use deaccentuation commonly embrace the Tokyo-centric gendered ideology portrayed by the mass media. Their use of the deaccented form is symbolic of their participation in this ideology.

Chapter 6 explored meanings of breathy voice and realization of /s/ by locating the variants in the participants’ stylistic practices. Following Eckert (1996), Bucholtz (1999), and Johnstone (1995), I defined style as a cluster of linguistic and non-linguistic features that are interpreted and understood by others. Also following Irvine’s (2001) claim, I examined individual styles in comparison with other people’s styles. Analyses of individual styles revealed that breathy voice helps construct mature femininity, which is aligned with Starr’s (2010) finding. I also illuminated how the mature female figure can be further categorized into two specific feminine figures based on whether the breathy voice is used with high pitch or low pitch. When breathy voice occurs with high pitch, the sensitive and traditional mature feminine figure is constructed, but when breathy voice occurs with low pitch, the strong and bold mature feminine figure is constructed. The analyses of individual styles also enhanced my argument that high frequency energy of /s/ is associated with sharpness in addition to the regional indexicality of the feature. Participants who display a higher COG of /s/ commonly present
attributes that are linked to sharpness such as thoughtlessness, competitiveness, and aggression, while participants who show lower COG of /s/ present more peaceful, modest, and unconfident figures.

Furthermore, Chapter 6 showcased how regional women negotiate the images their native language varieties carry by deploying a range of linguistic resources available to them. For example, a Kyoto participant, Chie, uses a low degree of breathy voice and a high COG of /s/ along with a wide pitch range and fast speech rate, which represents her aggressive style and contrasts with the stereotypical image of the Kyoto variety. An Osaka participant, Naho, avoids fronting /s/ unlike many other Osaka participants and uses a high degree of breathy voice, which constitutes the traditionally feminine style, rather than the harsh Osaka Auntie style. In contrast, Mariyo from Osaka aligns herself with Osaka Auntie while using a low degree of breathy voice and a high COG of /s/.

7.3. Contributions

The major contributions of this dissertation lie in the field of Japanese Dialectology and the field of Japanese Women’s Language. This dissertation problematizes previous work in the field of Japanese Dialectology. While Kyoto and Osaka are commonly located within the same dialect region, the analyses of this study show otherwise by presenting distinctions between the two regions. Chapter 3 demonstrates that breathy voice is used differently in Kyoto and Osaka, in addition to the difference in use in Kyoto and Tokyo. Chapter 4 reveals that while the realization of the fricatives /s/ and /ɕ/ in Osaka are significantly different from Tokyo, realization of the fricatives in Kyoto do not vary from Tokyo. Chapter 5 shows that the Tokyo-dominant intonational pattern, the deaccented form of negative questions, is used significantly more frequently in Kyoto than in Osaka. These findings reveal that the traditional dialectal divisions
created based on lexical and grammatical variation are insufficient to capture regional variation as well as local identities.

While studies of Japanese Dialectology have not paid much attention to co-occurring features of dialectal variants, this study shows the importance of considering how linguistic features work together in the construction of regional varieties. As is shown in Chapter 3, the interaction between breathy voice and intensity as well as the interaction between breathy voice and Fo exhibit significantly different patterns between Kyoto and the other two regions. This demonstrates that examinations of co-occurrence patterns for linguistic features are beneficial in the studies of Japanese Dialectology.

On the other hand, while previous work on Japanese Women's Language assumes a monolithic speech variety for all Japanese women, this dissertation shows that women in Tokyo, Kyoto, and Osaka exhibit strikingly different speech patterns. Rather than constructing a uniform gender identity, Japanese women produce gendered figures that typify particular geographic regions while negotiating the regional stereotypes by converging with and diverging from them. In addition to homogenizing diversity of speech varieties, research on Japanese Women’s Language is based on the premise of a sole language ideology shared across different regions. However, this study has shed light on the importance of considering locally specific language ideologies by spotlighting variation in indexical meanings of a given feature across the regions.

Finally, this dissertation advocates for more cross-field research across the studies of Japanese Dialectology and Japanese Women’s Language. The findings of this study demonstrate that region and gender are not detachable from each other. Region is one of the orthogonal dimensions that is intricately intertwined into people’s gender identities. Features that present cross-regional variation simultaneously connote social meanings that are relevant to gender.
7.4. Limitations and Future Studies

One of limitations of this dissertation is concerned with the definition of region. This study has used the prefectural units for the regions of Kyoto and Osaka and additionally included the two adjacent cities, Yokohama and Kanagawa, into the Tokyo region for the convenience of data collection. These categorizations are legitimated based on the conventional dialectal divisions. Nevertheless, it is also true that regional variation can be observed within the prefectural units. Particularly, it is a commonly accepted notion in Osaka that the Osaka Prefecture can be divided into three sub-dialectal areas: Northern Osaka (where the Settsu dialect is spoken), Eastern Osaka (where the Kawachi dialect is spoken), and Southwestern Osaka (where the Senshu dialect is spoken). Variation across these three sub-dialects can be seen in morphological, lexical, and phonetic features (Umegaki 1962). Since the samples in each region of this study are not large enough to examine patterns across smaller divisions under each region’s prefectural unit, I did not consider variation by regional sub-division. However, for more comprehensive and locally specific analysis of regional variation, it is important to consider even smaller geographic definitions of speech communities.

The second limitation to be noted is that this dissertation lacks a discussion of socioeconomic class as a result of the strict Japanese IRB regulations that prohibit questions identifying the participants’ socioeconomic class (For the details, see Chapter 2), even though socioeconomic class has been considered a strong predictor of sociolinguistic variation. All the participants are students of private colleges, which places them into the category of a certain class. In particular, the tuition for private colleges is usually more expensive than tuition at national or public institutions, and scholarships are not widely awarded to Japanese students. These factors may lead us to assume that class is controlled among the participants. However,
from the fieldwork as well as the recorded conversations, I realized that class was not as unified as I thought that it would be. For example, some participants of my study are daughters of a medical doctor or a company president, while some others strive to pay for their own tuition because, as they explain, their families cannot afford it. This raises an important question for social scientists about how to cope with the norms or taboos of the communities of the study and the goal to uncover more “real” human practices.

Sociolinguistic research (e.g. Bell 1984, Coupland 1980, Schilling-Estes 1998, 2004) has provided much evidence for people’s performative style shifting based on factors such as personal identity, interpersonal and group alignment, topics/alignment to topic, and speech genre. My study was originally designed to examine intraspeaker variation across conversational topics like women’s fashion, men, and careers, which was one of the reasons why I used different sets of magazines for the conversation recording. However, my focus on obtaining more relaxed conversations resulted in inconsistency of the conversational topics. Since I specifically instructed them to feel free to talk about any topics, a few pairs of the participants did not use the magazines while being recorded. Even in the conversations among the participants who used the magazines, it is very difficult to categorize topics using a consistent coding scheme.

Though topic-based stylistic analyses may not be very compatible with my data, it would be interesting to investigate the conversations as dyadic and how much each participant’s speech works to shape the other’s in a future work. As we saw in Chapter 5, four out of the five Kyoto participants who use the deaccented form of negative questions consisted of two pairs of conversational partners who both used it together. From my interspeaker analysis, it is not clear if the convergence is due to speech accommodation within the pairs or the particular communities of practice (or both). This point is more likely to become clear in intraspeaker analysis of each speaker’s speech. Intraspeaker analysis is also beneficial to investigate the
participants’ active identity work. It was not possible in my interspeaker analysis to distinguish, for example, breathy voice caused by the speakers’ physiology, from breathy voice agentively used as a stylistic resource.

Also, this study focuses mainly on production aspects of the phonetic variables and leaves perceptual perspectives out of the analysis. In the field of sociolinguistics, the vast majority of previous research has concentrated on speech production, and speech perception has remained relatively neglected (Thomas 2002), but there has recently been increasing interest in the perception approach due to recognition of the fact that perceptual mechanisms are needed to account for language variation (Labov 2006). In my study, perceptual experiments should enhance the discussions of social meanings of the variants and help verify that the acoustic measures of breathy voice and intensity or Fo, and the spectral moments of the fricatives actually do correspond with perceptions of breathiness, sharpness, clarity, and intensity. The notion that different feminine images are constructed through different patterns of co-occurrence of intensity, Fo, and degree of breathiness discussed in Chapter 3 can be corroborated by perception experiments. In Chapter 4, though I borrowed some psychoacoustic notions of sound qualities to interpret the social meaning of variation in the four spectral moments of the fricatives in the perceptual sense, a future study should investigate perceptual evaluations of sound sharpness or clarity as well as gendered images constructed by the different sound qualities by manipulating the four spectral moments of the fricatives.

As a next step to expand my work on sociophonetics and gender in Japanese, I aim to investigate articulation of the fricative /ɕ/ using media data. Media representations are a fruitful locus of investigation of social meanings of language variation since transparency of the meanings is the key to being rapidly consumed by a diverse viewership in media (Kajino and Podesva 2007). Additionally, media not only reflect but also shape both language use and ideologies about the nature of language as well as about the people who produce, consume
and/or are represented by media texts (Bell 1991, 1995, 2013; Coupland 2001, 2003; Jaffe 2011). Whereas my study has found that the Tokyo fricative /ɕ/ exhibits a significantly lower frequency energy than the Osaka fricative, I speculate that a few of the Tokyo speakers round their lips during articulation of the fricative just like the English /ʃ/, which results in a lower COG. To observe the states of speakers’ lips, visual data is required, which was not part of my study design. One hypothesis regarding the meaning of the lip rounding is that it connotes a certain type of metropolitan and global/Westernized personae. For example, the Tokyo speaker Aika, who I suspect exhibits this feature, presents orientations toward Western and, more specifically, American cultures. She works at a branch of the American restaurant Hooters in Tokyo and entered the Miss Universe Japan competition. She wants to become a Victoria’s Secret model and only shops for Western brands like Abercrombie & Fitch and Forever 21. Her last boyfriend is Japanese-Brazilian, and she claims that she has no interest in Japanese men. In contrast with many of the other participants who strive to obtain skinny and slender bodies and lighter and translucent skins, Aika desires to have a curvy and glamorous body with some muscles and uses a darker color makeup foundation. While Westernized/Americanized personae have been a target of strong hatred and ridicule and have not been accepted as mainstream for a long time in Japan, this feature may highlight a new Japanese gendered figure that has become increasingly visible in today’s Japanese society. This new feminine figure also contrasts with the ultra-feminine burikko figure, which is today regarded only in a deprecatory sense and is reportedly represented by high frequency energy of the fricatives according to some research (F. Inoue and Ogino 1984, 1985; cited in Inoue 1989) in the 80s (see Section 4.3.6 in Chapter 4).

While it is my best hope to submit this study toward the betterment of the participants’ lives, I plan to continue to search for more ways to give back to the community. During the fieldwork for this study, I met many students who were interested in study abroad or careers using English and offered advice, which was a way to give back to the community. I also gave a
talk to the entire student body of the English department of one research site about my cross-cultural experiences in the United States. For those participants who shared serious concerns (e.g. bullying, sexual harassment) and thoughts with me, it is my objective to continue exploring and writing about what I, as a sociolinguist, can do to create awareness and appreciation for the vast diversity of individual identities.

Although many issues still remain to be explored, this study has attempted to uncover the complex interplay of gender and region in Japanese women’s speech through examinations of less studied phonetic features. The quantitative patterning of breathy voice, acoustic characteristics of fricatives, and intonation across the regions have been linked to the intra-regional and individual stylistic practices in the dynamic sociolinguistic sphere. In closing, I hope that this dissertation has successfully presented theoretical and methodological approaches to help advance our understanding of Japanese women’s linguistic practices.
# Appendix I: Participants List

<table>
<thead>
<tr>
<th>Region_no</th>
<th>Speaker</th>
<th>Age</th>
<th>Year</th>
<th>Department</th>
<th>Parents’ home prefecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokyo_1</td>
<td>Aika</td>
<td>20</td>
<td>junior</td>
<td>English</td>
<td>Tokyo</td>
</tr>
<tr>
<td>Tokyo_2</td>
<td>Akiko</td>
<td>20</td>
<td>junior</td>
<td>English</td>
<td>Okayama</td>
</tr>
<tr>
<td>Tokyo_3</td>
<td>Yasuko</td>
<td>19</td>
<td>freshman</td>
<td>Japanese</td>
<td>Kagoshima</td>
</tr>
<tr>
<td>Tokyo_4</td>
<td>Chami</td>
<td>19</td>
<td>freshman</td>
<td>English</td>
<td>Kanagawa</td>
</tr>
<tr>
<td>Tokyo_5</td>
<td>Fuyu</td>
<td>21</td>
<td>junior</td>
<td>Social Science</td>
<td>Saitama</td>
</tr>
<tr>
<td>Tokyo_6</td>
<td>Hana</td>
<td>19</td>
<td>freshman</td>
<td>Psychology</td>
<td>Ibaraki</td>
</tr>
<tr>
<td>Tokyo_7</td>
<td>Takako</td>
<td>18</td>
<td>freshman</td>
<td>Japanese</td>
<td>Shizuoka</td>
</tr>
<tr>
<td>Tokyo_8</td>
<td>Jun</td>
<td>21</td>
<td>junior</td>
<td>English</td>
<td>Tokyo</td>
</tr>
<tr>
<td>Tokyo_9</td>
<td>Mei</td>
<td>21</td>
<td>junior</td>
<td>English</td>
<td>Tokyo</td>
</tr>
<tr>
<td>Tokyo_10</td>
<td>Megu</td>
<td>20</td>
<td>freshman</td>
<td>English</td>
<td>Tokyo</td>
</tr>
<tr>
<td>Tokyo_11</td>
<td>Ruka</td>
<td>19</td>
<td>freshman</td>
<td>Social Science</td>
<td>Tokyo</td>
</tr>
<tr>
<td>Tokyo_12</td>
<td>Rumiko</td>
<td>20</td>
<td>junior</td>
<td>English</td>
<td>Tokyo</td>
</tr>
<tr>
<td>Tokyo_13</td>
<td>Yoshiko</td>
<td>21</td>
<td>senior</td>
<td>English</td>
<td>Tokyo</td>
</tr>
<tr>
<td>Tokyo_14</td>
<td>Kasumi</td>
<td>20</td>
<td>junior</td>
<td>Social Science</td>
<td>Osaka</td>
</tr>
<tr>
<td>Tokyo_15</td>
<td>Yuka</td>
<td>18</td>
<td>freshman</td>
<td>English</td>
<td>Aomori</td>
</tr>
<tr>
<td>Kyoto_1</td>
<td>Akari</td>
<td>18</td>
<td>freshman</td>
<td>Japanese</td>
<td>Kyoto</td>
</tr>
<tr>
<td>Kyoto_2</td>
<td>Nao</td>
<td>20</td>
<td>junior</td>
<td>English</td>
<td>Kyoto</td>
</tr>
<tr>
<td>Kyoto_3</td>
<td>Chie</td>
<td>20</td>
<td>junior</td>
<td>English</td>
<td>Hiroshima</td>
</tr>
<tr>
<td>Kyoto_4</td>
<td>Futaba</td>
<td>21</td>
<td>senior</td>
<td>English</td>
<td>Gifu</td>
</tr>
<tr>
<td>Kyoto_5</td>
<td>Mami</td>
<td>20</td>
<td>senior</td>
<td>English</td>
<td>Kyoto</td>
</tr>
<tr>
<td>Kyoto_6</td>
<td>Mika</td>
<td>18</td>
<td>freshman</td>
<td>English</td>
<td>Kyoto</td>
</tr>
<tr>
<td>Kyoto_7</td>
<td>Sayumi</td>
<td>18</td>
<td>freshman</td>
<td>Japanese</td>
<td>Shiga</td>
</tr>
<tr>
<td>Kyoto_8</td>
<td>Mitsuki</td>
<td>18</td>
<td>freshman</td>
<td>Japanese</td>
<td>Aichi</td>
</tr>
<tr>
<td>Kyoto_9</td>
<td>Sakura</td>
<td>20</td>
<td>junior</td>
<td>English</td>
<td>Okayama</td>
</tr>
<tr>
<td>Kyoto_10</td>
<td>Eri</td>
<td>18</td>
<td>freshman</td>
<td>English</td>
<td>Osaka</td>
</tr>
<tr>
<td>Kyoto_11</td>
<td>Rika</td>
<td>18</td>
<td>freshman</td>
<td>Japanese</td>
<td>Kyoto</td>
</tr>
<tr>
<td>Kyoto_12</td>
<td>Orie</td>
<td>21</td>
<td>senior</td>
<td>English</td>
<td>Osaka</td>
</tr>
<tr>
<td>Kyoto_13</td>
<td>Tomomi</td>
<td>21</td>
<td>junior</td>
<td>English</td>
<td>Kyoto</td>
</tr>
<tr>
<td>Kyoto_14</td>
<td>Sae</td>
<td>21</td>
<td>senior</td>
<td>English</td>
<td>Kyoto</td>
</tr>
<tr>
<td>Kyoto_15</td>
<td>Yumi</td>
<td>21</td>
<td>senior</td>
<td>English</td>
<td>Kyoto</td>
</tr>
<tr>
<td>Osaka_1</td>
<td>Ana</td>
<td>18</td>
<td>freshman</td>
<td>Nursing</td>
<td>Osaka</td>
</tr>
<tr>
<td>Osaka_2</td>
<td>Asako</td>
<td>18</td>
<td>freshman</td>
<td>English</td>
<td>Osaka</td>
</tr>
<tr>
<td>Osaka_3</td>
<td>Reina</td>
<td>18</td>
<td>freshman</td>
<td>Nursing</td>
<td>Kagawa</td>
</tr>
<tr>
<td>Osaka_4</td>
<td>Kana</td>
<td>21</td>
<td>senior</td>
<td>Psychology</td>
<td>Osaka</td>
</tr>
<tr>
<td>Osaka_5</td>
<td>Taeko</td>
<td>20</td>
<td>junior</td>
<td>Japanese</td>
<td>Osaka</td>
</tr>
<tr>
<td>Osaka_6</td>
<td>Saori</td>
<td>20</td>
<td>junior</td>
<td>Japanese</td>
<td>Osaka</td>
</tr>
<tr>
<td>Osaka_7</td>
<td>Midori</td>
<td>18</td>
<td>freshman</td>
<td>Nursing</td>
<td>Osaka</td>
</tr>
<tr>
<td>Osaka_8</td>
<td>Naho</td>
<td>19</td>
<td>sophomore</td>
<td>English</td>
<td>Fukuoka</td>
</tr>
<tr>
<td>Osaka_9</td>
<td>Miya</td>
<td>21</td>
<td>senior</td>
<td>Psychology</td>
<td>Osaka</td>
</tr>
<tr>
<td>Osaka_10</td>
<td>Kimi</td>
<td>19</td>
<td>sophomore</td>
<td>English</td>
<td>Hiroshima</td>
</tr>
<tr>
<td>Osaka_11</td>
<td>Nanako</td>
<td>18</td>
<td>freshman</td>
<td>Nursing</td>
<td>Osaka</td>
</tr>
<tr>
<td>Osaka_12</td>
<td>Mariyo</td>
<td>19</td>
<td>sophomore</td>
<td>English</td>
<td>Saga</td>
</tr>
<tr>
<td>Osaka_13</td>
<td>Saya</td>
<td>18</td>
<td>freshman</td>
<td>Nursing</td>
<td>Osaka</td>
</tr>
<tr>
<td>Osaka_14</td>
<td>Kei</td>
<td>18</td>
<td>freshman</td>
<td>Nursing</td>
<td>Hyogo</td>
</tr>
<tr>
<td>Osaka_15</td>
<td>Teru</td>
<td>19</td>
<td>sophomore</td>
<td>English</td>
<td>Osaka</td>
</tr>
<tr>
<td>Osaka_16</td>
<td>Yuriko</td>
<td>18</td>
<td>freshman</td>
<td>English</td>
<td>Osaka</td>
</tr>
</tbody>
</table>
Appendix II: Interview Questionnaire

<table>
<thead>
<tr>
<th>Background information</th>
</tr>
</thead>
<tbody>
<tr>
<td>- How old are you?</td>
</tr>
<tr>
<td>- In which prefecture/city did you grow up? (only for the Tokyo participants to identify participants from Tokyo, Yokohama, and Kawasaki)</td>
</tr>
<tr>
<td>- In which prefecture/city do you currently live?</td>
</tr>
<tr>
<td>- From which prefectures are your parents are?</td>
</tr>
<tr>
<td>- How long have you been friends with your conversational partner?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Community of practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Do you work? Where do you work?</td>
</tr>
<tr>
<td>- Do you belong to any club activities?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attitude toward their universities</th>
</tr>
</thead>
<tbody>
<tr>
<td>- What do you study?</td>
</tr>
<tr>
<td>- What made you choose your university? What made you choose a women’s university?</td>
</tr>
<tr>
<td>- What do you think about your university?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attitude toward their regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Do you like Kyoto/Osaka/Tokyo? Why/why not?</td>
</tr>
<tr>
<td>- Do you plan to stay in Kyoto/Osaka/Tokyo after graduation? Why?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attitude toward their varieties</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Do you think you speak the Kyoto/Osaka/Tokyo variety?</td>
</tr>
<tr>
<td>- Are you proud of speaking your own variety? Why/why not?</td>
</tr>
<tr>
<td>- Have you had anyone say anything positive/negative about your variety?</td>
</tr>
<tr>
<td>- Did your parents say anything about the way you speak when you were a child or recently?</td>
</tr>
<tr>
<td>- What kind of images do you have toward the Kyoto, Osaka, and Tokyo varieties?</td>
</tr>
<tr>
<td>- Among the three varieties, which do you think is the most feminine/masculine?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attitude toward Japanese Women’s Language (JWL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Do you think you use JWL in daily life? If not, why don’t you use it? If yes, do you use it all the time?</td>
</tr>
<tr>
<td>- If anything, in which situations do you use it?</td>
</tr>
<tr>
<td>- What kind of images do you have toward JWL or women who use JWL?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>- What is your favorite fashion brand? Where do you usually shop for your clothes?</td>
</tr>
<tr>
<td>- Can you tell me what your email address means?</td>
</tr>
</tbody>
</table>
Appendix III: Original Japanese Transcriptions of Excerpts in Chapter 6

Excerpt 6.1 (p.225-26)
1. Jun: えー…男っぽすぎるとか。あと．まあ男っぽすぎる
   Well… Like I sound too manly. And.. yea, too manly.
2. Int: なんで男っぽいんだろ？
   I wonder what makes [your speech] sound manly..
3. Jun: 普段なんか「オレ」とか言ってました
   I used to say *ore*
4. Int: へー、なんで「オレ」って言ったの？
   Mm hmm, why were you using *ore*?
5. Jun: 女がイヤで…。
   I didn’t like being a girl...
6. Int: あー、そうなんだ。
   I see.
7. Jun: 一時期、「なんで女なんだろ？」
   For awhile, I was like “Why am I a girl?”
8. Int: え、何歳くらいの時？
   Eh, how old were you?
9. Jun: 中学の時。男がすっごくかっこ良く見えたんですよ。
   Junior high. Boys looked really cool to me.
10. Int: うん。
   Mm Hmm, yea.
11. Jun: で、まあ好きな人ができるにつれて、それはもう変わったんだけど、それまでは、
    Then, I had a crush, and it changed, but until then, boys looked simply
12. もう男がかっこよくて、こう、男って.. 何かしらなんかこう、スポーツやるにしても
    cool. And boys are like.. somehow, devoting 100% of their energy to
13. 勉強にしても、何事に対しても一生懸命突っ走る感があって、女の子って周りの目、
    playing sports, studying, or doing anything, but girls worry a lot about
14. すっごい気にするじゃないですか。それがいやで、「うーん男かっこいいな、
    how they are seen, you know? I hated it, so I was like “Oh my god, boys
15. なんで私男じゃないんだろ」みたいな。あと、周りの女の子達も、背が
    aren so cool, why am I not a boy?” Also since I was taller than other girls, I
16. おっきいかから、「かわいい」とか「きれい」とかより、「かっこいい
    was always described by other girls as “cool/handsome” rather than “cute”
17. って言われてた。
    or “pretty.”

Excerpt 6.2 (p. 227)
1. Mei: ありますね、でもなんか…なんだろ、あchatの場合は結構、その今もそ-なんか自分
   Yes, I have, but like.. well, for me, it’s not... really I don’t know why,
2. では結構、あんまり感じないんですけど、その…なんかちょっと偉い人って言うか年
   but well: .. before, when I met people above me or older people.; It was
3. 上の人に会った時に、「ちょっと君はしゃべり方が年齢に幼すぎてるって言うか、子
   said; “the way you speak sounds too immature relative to your age or
4. に供しつけ方するから直した方がいいよ」って言われたことがあるって、なんか
   you speak in a childish manner, so it’s better to change the way
5. <laugh>
you speak” <laugh>

6. Int: なんでだる。
Hm, I wonder why?

7. Jun: ゆっくり-ゆっくりだからじゃない？のろのろしゃべるから。
Isn’t it because you speak slowly? Because you drawl.

8. Mei: <laugh> あ、そうなのかな。でも全然私は分からないんですけど、なんかすごい..言う
<laugh> Ah I wonder if that’s the case. But I totally have no idea: like

9. くれます。
often.. people say that to me...

10. Jun: メリハリじゃない？
Because [your speech] lacks dynamism [in the rhythm]?

11. Mei: メリハリがないのかな <laugh>  RESET. なんか早くバラっと話すのできなくてー。
No dynamism <laugh>. But I can’t fix it:. I can’t speak fast:.

12. Jun: でも顔もあると思うよ。
But I think the impression has something to do with her looks.

13. Int: どんな風に？
How?

14. Jun: だからメイの顔で、そうしゃべるから「ゆっくり」で済むけど、でもこうゆう顔
Since Mei talks like that with her looks, it’s considered immature, but if [I]

15. <pointing Jun’s own face> でゆっくりしゃべったら、
talk like a kid with these looks <pointing to her own face>, then people

16. 多分「ブリッコ」ってなるよ。
would describe it as burikko.

Excerpt 6.3 (p. 230)

1. Kana: 週 5！
5 days a week!

2. Miya: やばない、アイツ。
Yea, she ‘aitsu’ is so crazy.

3. Kana: すごいな、ようやるな…。しかもそれで一回生やで。
Wow, how does [she] do that… Plus, [she] is still in her first year and

4. has many classes.

5. Miya: そうやで、朝一限多いし。
Yea, there are a lot of morning classes.

6. Kana: ようやっていける。うちだったら絶対無理やわ。
I can’t believe that [she] does that. I would totally fail at it.

7. Miya: アツ子、すごい。そこでチア好きなんやろな。
Atsuko is amazing. I guess [she] loves cheerleading that much.

8. Kana: 今の一回生ヤバない？
Aren’t all the freshmen this year tough?

9. Miya: バイトの量、ヤバない？
Yea, [they] all work like crazy.

10. Kana: それで、今全然普通にできてるからいいのかも知らんけど。
[They] are doing okay in training, so I guess that’s no problem.

11. Miya: あの子ら、なんか全然困る
They ‘anoko-ra’ are not getting to the point where [they] have to

12. 時代じゃないやん。
struggle yet.

13. Kana: まあねまあね。
   Yea yea.

14. Miya: なんでもできるやん。ばんばんばーん、みたいな。
   [They] can do anything now.

15. Kana: な。もうできてるもん、ほとんど。
   Yea, and [they] are already doing well.

16. Miya: うらやましいね。
   I miss those days.

17. Kana: ほんとね。
   Truly.

Excerpt 6.4 (p. 231-32)

1. Nana: ほんま、ナナ、就職せんと結婚しようかな。
   I <'Nana'> really think of getting married without getting a job.

2. Ana: うん、したい、あたしも。
   Yea, sounds good to me <'atashi'> too.

3. Nana: ほんまに、働かんと結婚するわ。
   Seriously, [I] will get married and not work.

4. Ana: でも、同じ年でな、なかなかおらんで。
   But it's so rare to find someone like that in our age, you know?

5. Nana: おらんな。絶対な、まあ男は、就職したとして、で、まあナナらは、
   Yea, so rare. Let's say, our boyfriends find jobs, and we <'Nana-ra'>

6. 普通に就職せんとその人と結婚したら、
   are married without getting jobs. And then, the places [we] live

7. 絶対自分の家ではないやん。多分親の家に居候やん、とりあえず、
   wouldn't be [our] own home, but [our] parents' home, until enough

8. お金貯まるまで。
   money is saved.

9. Ana: 同い年だったら絶対そうなるな。
   If the husbands are our age, that's the case.

10. Nana: それはいや。
    It sounds horrible.

11. Ana: いやや。
    Right.

12. Nana: 親の世話にはなりたくないわ。
    [I] don't want to ask for parents' help.

13. Ana: でも、専業主婦は絶対やらんと思う、
    But if it were like that, [I] don't think I would become a stay-home-

14. そんなんやったら。その状況やったらさ。
    wife for sure, not if the situation were like that.

15. Nana: ちょっと、働き出てもそんな大したお金･･･
    But even if [we] go work, [we] wouldn't make that much money...

16. Ana: あの、まあ、どうせパートになるかもしれないけど。
    Yea, maybe it would just be a part-time job.

Excerpt 6.5 (p. 234)
Mika: I've been wondering, I've been wondering for awhile, what do people do once they are in a relationship?.. Like, I've been thinking.

Rie: What an interesting person you are...

Mika: <laugh> Eh, because I was just wondering. Well, in romance, before you get serious.. it’s the best time. Somehow like ahhh..

Rie: Uh... when you have a crush, right?

Mika: Right, I think a crush is fun, like quite fun.

Rie: It’s exciting

Mika: think fighting isn’t bad.

Rie: If you are very passive like you can’t fight, and you keep everything to yourself, then maybe you shouldn’t be in the relationship. That’s how I broke up <laugh>.

Mika: Oh really?

Rie: すっごい、なんか、あんまり意見言えへんかった。

Excerpt 6.6.a (p. 237)

Tomo: あたし、中学時代、なんかその、お腹すいてたのに、いっぱい食べるってゆうの。When I was in junior high, somehow, even when I was hungry, I felt

が恥ずかしいって、家でも。なんか、こう、家でやっぱり弟もいたし、お母さんがembarrassed about eating a lot, even at home. Like at home.. since I have a
　が、おやつとかポテトチップスとか買ってきてはんねんけど、弟はあーって食べbrother, my mom was buying snacks like potato chips, and my brother wasってねんか、でも「お姉ちゃんも食べへんの？」みたいに言われて、言わ- (え、eating them really fast, but when he asked, “Aren’t you eating?” to me, (-yea,なに？）「あげるわ」とか言われても、なんか「いらん」とか言って...。た- あんまwhat?) So when he offered the snacks to me, I was like “no thanks”... I- Iり食べへんかった。ほんと人前でその... didn’t eat it. Because, in front other people, like...

Maya: 助ずかしいん？
Were you embarrassed?
8. Tomo: そう、そのいっぱい食べるのが。
    Yes, about eating a lot.

Excerpt 6.6 (p. 237)
1. Tomo: 電話でしゃべってて... ええ、なんかその... その笑いの反動で出ちゃった
   We were talking on the phone... and like... It just came out when I
   thought... laughed.
2. ときがあって。
3. Maya: うんうん、そんな全然気にせんでいいやる。
   I see.. You don’t have to worry about that at all.
4. Tomo: 電話やし、開こえんかなと思ってんけど...
   It was over the phone, so I thought he didn’t hear it... Do you think
   it was all right?
5. Maya: 大丈夫かな。
6. Tomo: 電話やし、開こえんかなと思ってんけど...
   It was over the phone, so I thought he didn’t hear it... Do you think
   it was all right?
7. Maya: 大丈夫大丈夫、気にしないよ。でも- でも、ケンさんやったら、しても笑ってくれ
   No problem, no problem. You worry too much. But-but even if Ken
   heard it, I feel like he would just laugh at it.
8. Tomo: いやいやいや、その気付かれんのがいや、私は。<laugh>
   No no no no, I just can’t stand the idea of him noticing it <laugh>.

Excerpt 6.7 (p. 243)
1. Mami: めっちゃ派手で、これ絶対合わへんやろって思ったシャツをな、アヤ子ちゃんと
   So Ayako brought me a shirt like “what about this?” which I thought
   な、これどうですかって言われて。
   wouldn’t look good on me at all because it looked so flashy
2. Nao: 意外と- 意外と見えてくるって。
   Unexpectedly- unexpectedly, you might see something.
3. Mami: そう。で、え？これ大丈夫？みたいになって言ったら、なんか、え、これ絶対大丈夫
   Yea, and I said like “Eh? Is this all right?” Ayako was like “yeea, this is
   です、みたいに言われて、着てみて、で、結局なんか店員さんに
   definitely good,” so I tried it on.. Then, eventually, the store clerk
4. Mami: そう。で、え？これ大丈夫？みたいになって言ったら、なんか、え、これ絶対大丈夫
   Yea, and I said like “Eh? Is this all right?” Ayako was like “yeea, this is
   です、みたいに言われて、着てみて、で、結局なんか店員さんに
   definitely good,” so I tried it on.. Then, eventually, the store clerk
5. Nao: 意外と- 意外と見えてくるって。
   Unexpectedly- unexpectedly, you might see something.
   なぜ啊はるやん、
   flattered me, like they give you lots of compliments, you know? Like
7. Nao: 意外と- 意外と見えてくるって。
   Unexpectedly- unexpectedly, you might see something.
   なぜ啊はるやん、
   flattered me, like they give you lots of compliments, you know? Like
   に二人にこうニコニコ見られて、のせられて買ってしました。<laugh>
   looked at me smiling, so I couldn’t say no... <laugh>

Excerpt 6.8 (p. 249)
   なぜ啊はるやん、
   flattered me, like they give you lots of compliments, you know? Like
   their ages, like middle-aged men and women in their 50s and 60s.
2. Nao: 意外と- 意外と見えてくるって。
   Unexpectedly- unexpectedly, you might see something.
   なぜ啊はるやん、
   flattered me, like they give you lots of compliments, you know? Like
   their ages, like middle-aged men and women in their 50s and 60s.
   絶対、年齢分かってない・・と思う。<laugh>
They always start talking to me! They totally think of my age wrong. I think <laugh>. In fact, it’s true that I am not in my 20s in my heart, definitely <laugh>. 

Int: マジで！？
Are you serious!?

Mari: 50-60のおっちゃんがしゃべってくるね。だから、店とか行ったたら、Yea, old men in their 50s and 60s start talking to me! And when I go to stores, they ask me like “What’s the great deal today?”, so I’m like “I’m not working here!” <laugh>. Even at a drug store, I was asked like “What’s on sale today?” so I-I think to myself like “I’m not a pharmacist!,” it’s everywhere, at grocery stores, some people asked me, “Which store has the best deal?” but somehow, I can answer those questions <laugh>.

Kimi: それで答えられるっていうのがすごい。<laugh>
It's amazing you can answer them <laugh>.

Mari: チラシを見て買いに行くからさ。<laugh>
Yea, I check out those store ads before going shopping <laugh>.
References:


Hirano, M. Clinical Examination of Voice. London: Springer.


the Acoustical Society of America 84: 511-529.


