CROSS-LANGUAGE STUDY OF AGE PERCEPTION

Kyoko Nagao

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Linguistics
and the Department of Speech and Hearing Sciences
Indiana University
Bloomington, Indiana
September, 2006
Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____________________________________
Kenneth de Jong, Ph.D.

_____________________________________
Diane Kewley-Port, Ph.D.

_____________________________________
Karen Forrest, Ph.D.

August 15, 2006

Robert Port, Ph.D.
This dissertation is dedicated

in loving memory of my grandmother

Fumi Nagao (1914-2006)

This dissertation is also dedicated to

my father, Yasunori Nagao
Acknowledgements

This dissertation was supported various people and organizations. I would like to thank Indiana University Graduate School, Indiana University Graduate and Professional Student Organization, Indiana University Department of Linguistics, and Indiana University Department of Speech and Hearing Sciences for enabling me to carry out both recording and perception experiments in Japan and the United States.

My deepest appreciation goes to Kenneth de Jong for his extraordinary mentorship since first coming to Indiana University as a master student. Without his academic guidance and constant encouragement, I could not have received even a Master’s degree. I also thank another supervisor Diane Kewley-Port for her helpful supervision. It was she who first welcomed me to the field of speech and hearing sciences. I could not pursue a dual doctoral degree in the two departments without her support. I also thank the two other members of my committee, Bob Port and Karen Forrest for valuable suggestions to this project.

I also thank Haruo Kubozono (Kobe University) for allowing me to use his facility for the recording in Japan and Miho Mano (Junshinkai Kango Iryoo University) and members of Buyookai Hyogo High School Alumni Association (in Hyogo, Japan) for their assistance in locating Japanese participants. I thank Miyako Muto (Kobe Women’s University) for her generous help for the perception experiment in Japan.

I also thank current and former IU fellows for their constant encouragement.
Finally, I thank my family, my father Yasunori Nagao, my grandmother Fumi Nagao, my aunt Yoshiko Haraguchi, and my cats Minton and Tawara, for all their love and support over the years. Especially, I thank my grandmother who inspired this project and brought me a whole new perspective of speech science.
ABSTRACT

CROSS-LANGUAGE STUDY OF AGE PERCEPTION

A number of studies have shown that listeners can estimate the age of talkers quite accurately by listening to speech alone. However, the effects of native language on age perception have not yet been explored. The current study examined the effects of listener’s language familiarity on the perception of a talker’s age in the three linguistic contexts varying the amount of information, i.e., vowel, phrase, and sentence. Two groups of listeners (English and Japanese) estimated the age of talkers whose native language were matched or mismatched with the listener’s. Furthermore, in order to investigate the effect of age stereotypes in each language, the same listeners estimated the age of talkers who disguised themselves as 20 year older or younger than their age.

Results indicated that listener’s estimation of talker’s age improved when more information was available. The listeners estimated the age of talkers more accurately in the familiar language than the foreign language. Better age estimation was found for female talkers than male talkers, but the effect of talker’s sex only appeared in the age estimation in the familiar language. Results of age estimation for age-disguised speech revealed that both language groups in this investigation have similar age stereotypes. These results suggested that the age-related speech characteristics are based on both on physiological factors and linguistic variation, variation that a non-native listener does not have access to. Results also suggest that there exists an underlying perceptual mechanism for identifying the age that is common across languages.
# Table of Contents

ACKNOWLEDGEMENTS ........................................................................................................ V

ABSTRACT ....................................................................................................................... VII

TABLE OF CONTENTS ................................................................................................. IX

LIST OF TABLES .......................................................................................................... XI

LIST OF FIGURES ..................................................................................................... XIV

1. INTRODUCTION AND LITERATURE REVIEW ...................................................... 1
   - INTRODUCTION .............................................................................................................. 1
   - AGE PERCEPTION RESEARCH .................................................................................... 2
   - AGE-RELATED ACOUSTIC CHARACTERISTICS OF VOICE AND SPEECH .......... 4
   - PERCEPTUAL AND ACOUSTIC CORRELATES OF AGE PERCEPTION ............... 12
   - OTHER FACTORS INFLUENCING AGE PERCEPTION ............................................. 15
   - SUMMARY ................................................................................................................... 32

2. SPEECH CORPUS .................................................................................................. 34
   - PARTICIPANTS ............................................................................................................. 35
   - SPEECH MATERIALS ................................................................................................. 39
   - RECORDING PROCEDURES ....................................................................................... 42

3. THE EFFECTS OF LANGUAGE FAMILIARITY AND THE AMOUNT OF INFORMATION ........................................ 49
   - INTRODUCTION .......................................................................................................... 49
   - PURPOSE ...................................................................................................................... 55
   - METHODS ................................................................................................................... 58
   - RESULTS ..................................................................................................................... 64
   - DISCUSSION ................................................................................................................ 77
   - CONCLUSIONS .......................................................................................................... 81

4. THE EFFECTS OF TALKER’S SEX .................................................................. 83
   - INTRODUCTION .......................................................................................................... 83
   - PURPOSE ...................................................................................................................... 88
   - METHODS ................................................................................................................... 90
   - RESULTS ..................................................................................................................... 91
   - DISCUSSION ................................................................................................................ 101
   - CONCLUSIONS .......................................................................................................... 105

5. THE EFFECTS OF AGE DISGUISE .................................................................. 106
   - INTRODUCTION .......................................................................................................... 106
   - METHODS ................................................................................................................... 113
   - RESULTS ..................................................................................................................... 116
   - DISCUSSION ................................................................................................................ 129
CONCLUSIONS .................................................................................................................... 135

6. GENERAL DISCUSSION AND CONCLUSION ..................................................... 136

SUMMARY OF RESULTS ................................................................................................. 136
PERCEPTUAL MECHANISMS IN IDENTIFICATION OF TALKER’S AGE ....................... 138
Exemplar-based model .................................................................................................. 138
Prototype-based model ............................................................................................... 139
Anchor model ............................................................................................................... 144
Abstractionist model .................................................................................................... 145
Evaluations .................................................................................................................... 146

LANGUAGE SPECIFIC ACOUSTIC CHARACTERISTICS ................................................. 152
FUTURE RESEARCH ..................................................................................................... 155
CONCLUSION ................................................................................................................ 158

REFERENCES ................................................................................................................. 160

APPENDIX A1: INFORMATION OF ENGLISH TALKERS ........................................ 177
APPENDIX A2: INFORMATION OF JAPANESE TALKERS ......................................... 180
APPENDIX B1: QUESTIONNAIRE FOR A TALKER (ENGLISH VERSION) ............... 183
APPENDIX B2: QUESTIONNAIRE FOR A TALKER (JAPANESE VERSION) .............. 189
APPENDIX C1: RECORDING INSTRUCTIONS (ENGLISH VERSION) ..................... 196
APPENDIX C2: RECORDING INSTRUCTIONS (JAPANESE VERSION) ....................... 207
APPENDIX D1: LISTENER QUESTIONNAIRE (ENGLISH VERSION) ....................... 219
APPENDIX D2: LISTENER QUESTIONNAIRE (JAPANESE VERSION) ....................... 222
CURRICULUM VITAE
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Types of the stimuli used in previous studies on age perception</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>Age perception studies in German, Italian, Japanese, and American English</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>Speakers' age range in the previous studies</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>Mean or Median correlation coefficients for female and male speakers</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>Mean and range of age of English speakers</td>
<td>36</td>
</tr>
<tr>
<td>6</td>
<td>Average years of smoking for English present/former smokers</td>
<td>37</td>
</tr>
<tr>
<td>7</td>
<td>Mean and range of age of Japanese speakers</td>
<td>38</td>
</tr>
<tr>
<td>8</td>
<td>Average years of smoking for Japanese present and former smokers</td>
<td>39</td>
</tr>
<tr>
<td>9</td>
<td>Summary of recording conditions and their order</td>
<td>44</td>
</tr>
<tr>
<td>10</td>
<td>Correlation values between perceived age and chronological age in various studies</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>Correlation values between perceived age and chronological age in various studies</td>
<td>54</td>
</tr>
<tr>
<td>12</td>
<td>Mean age and age range of English listeners</td>
<td>59</td>
</tr>
<tr>
<td>13</td>
<td>Mean age and age range of Japanese listeners</td>
<td>59</td>
</tr>
<tr>
<td>14</td>
<td>Order of stimulus presentation</td>
<td>62</td>
</tr>
<tr>
<td>15</td>
<td>Pearson correlations between PA and CA in terms of language familiarity</td>
<td>64</td>
</tr>
<tr>
<td>16</td>
<td>Pearson correlation coefficients between talkers’ chronological age and perceived age in Vowel, Phrase, and Sentence contexts</td>
<td>65</td>
</tr>
<tr>
<td>17</td>
<td>Analysis of Variance for Accuracy</td>
<td>69</td>
</tr>
<tr>
<td>18</td>
<td>ANOVA results for main effects and interaction effects of context, language familiarity, and listener language on age estimation accuracy for each age group of talkers</td>
<td>72</td>
</tr>
</tbody>
</table>
Table 19. Pearson correlation coefficients between CA and PA in each context based on the data from female listeners ............................................................................... 80

Table 20. Correlation coefficients for male and female speech obtained from various studies. .................................................................................................................. 87

Table 21. Predicted sex differences in the listener’s performance on age perception based on various accounts for English talkers only ......................................................... 88

Table 22. Mean and range of age of 30 English speakers and 30 Japanese speakers ...... 90

Table 23. Correlation between CA and PA for English talkers ........................................... 92

Table 24. Means and standard deviations of PA for English male and female talkers (young, middle-aged, and elderly) by English listeners ........................................... 93

Table 25. Correlation between CA and PA for English male and female talkers by English male and female listeners ............................................................................... 94

Table 26. Correlations between CA and PA for Japanese talkers ........................................ 98

Table 27. Means and standard deviations of PA for Japanese male and female talkers (young, middle-aged, and elderly) by Japanese listeners ....................................... 98

Table 28. Means and Standard Deviations of ∆age calculated as normal age minus rejuvenation age in years. ................................................................................... 116

Table 29. Two-way ANOVAs on perceived age estimated by English and Japanese listeners (L) for talkers in the rejuvenation and normal conditions (C). Four separate ANOVAs are shown for the English middle-aged and elderly talker groups, and Japanese middle-aged and elderly talker groups ......................................................... 117

Table 30. Average perceived age in the rejuvenation and normal conditions in years. 118

Table 31. Results of two separate three-way ANOVAs for listener’s language (L), talker’s sex (S), talker’s age group (A) on ∆age (= Perceived age in rejuvenation condition − Perceived age in normal condition) for each talker .... 119

Table 32. Means and Standard Deviations of ∆age calculated as maturation age minus normal age in years. ..................................................................................... 123

Table 33. Two-way ANOVAs on perceived age estimated by English and Japanese listeners (L) for talkers in the rejuvenation and normal conditions (C). Four separate ANOVAs are shown for English young and middle-aged talker groups, and Japanese young and middle-aged talker groups ........................................... 124
Table 34. Comparison of mean perceived age between the maturation and normal conditions in years. ................................................................. 125

Table 35. Results of two separate three-way ANOVAs for listener’s language (L), talker’s sex (S), and talker’s age group (A) on Δage (= Perceived age in maturation condition − Perceived age in normal condition) for each talker...... 128
List of Figures

Figure 1. Age estimation errors (=DA) for each age group by familiar vs. less-familiar listener groups in the vowel context. .................................................................67

Figure 2. Age estimation errors (=DA) for each age group by familiar vs. less-familiar listener groups in the phrase context. .................................................................67

Figure 3. Age estimation errors (=DA) for each age group by familiar vs. less-familiar listener groups in the sentence context. .................................................................68

Figure 4. Accuracy of age estimation in terms of listener language, language familiarity for young talkers in the phrase and sentence contexts. .......................72

Figure 5. Accuracy of age estimation in terms of listener language, language familiarity for middle-aged talkers in the phrase and sentence contexts. ..........73

Figure 6. Accuracy of age estimation in terms of listener language, language familiarity for elderly talkers in the phrase and sentence contexts. ...............73

Figure 7. Average magnitude of age estimation errors (=|DA|) in years for the talkers of a familiar language in each context by each listener group. .................74

Figure 8. Average magnitude of age estimation errors (=|DA|) in years for all talkers of a less-familiar language in each context by each listener group. .................75

Figure 9. Difference between perceived age and chronological age (DA) in terms of talkers’ sex and age groups for the English sentence stimuli by English listeners. DA for male talkers is indicated by squares and for female talkers by triangles.................................................................95

Figure 10. Difference between perceived age and chronological age (DA) in terms of talkers’ sex and age groups for the English sentence stimuli by Japanese listeners. DA for male talkers is indicated by squares and for female talkers by triangles.................................................................96

Figure 11. Difference between perceived age and chronological age (DA) in terms of talkers’ sex and age group for the Japanese sentence stimuli by English listeners. DA for male talkers is indicated by squares and for female talkers by triangles.................................................................99

Figure 12. Difference between perceived age and chronological age (DA) in terms of talkers’ sex and age group for the Japanese sentence stimuli by Japanese
listeners. DA for male talkers is indicated by squares and for female talkers by
triangles................................................................................................................100

Figure 13. Differences in the perceived age for English talkers between the
rejuvenation and normal conditions.................................................................120

Figure 14. Differences in the perceived age for Japanese talkers between the
rejuvenation and normal conditions.................................................................121

Figure 15. Differences in the perceived age for English talkers between the
maturation and normal conditions. ..................................................................127

Figure 16. Differences in the perceived age for Japanese talkers between the
maturation and normal conditions. ..................................................................128

Figure 17. Age perception in a prototype-based model...........................................141

Figure 18. Pitch contours of the phrase stimulus by English middle-aged woman,
EFM07 (Top panel) and Japanese middle-aged woman, JFM04 (Bottom
panel) in the normal condition........................................................................154
CHAPTER 1

1. Introduction and Literature Review

Introduction

While linguistic studies on speech have been based on speech samples collected from various age ranges of speakers, speech by elderly people has been the least studied area in linguistics. Most of the studies on elderly speech production and perception or speech production and perception related to advancing aging are in the field of speech-language pathology and audiology. These studies are more or less clinically oriented since speech-language pathologists need to properly diagnose the patients for appropriate clinical intervention. For elderly speakers, some symptoms of communicative disorders are purely pathological, but they could also be a part of common aging process. However, most of the clinically oriented studies tend to focus on disordered populations. General speech characteristics of the elderly people are studied for collecting normative data to compare with age-matched disordered speech data. In spite of the importance of normative data on non-pathological elderly speech, there is very sparse literature about general speech characteristics of the normal elderly speakers.

More than a decade has passed since Battle emphasized the need for multicultural perspectives in the field of speech pathology (Battle, 1993). Still, it does not appear that ethnic differences have been well incorporated with basic geriatric speech research. In spite of the fact that there is a growing number of people with different cultural and
linguistic backgrounds in the United States, the majority of research is based on English (usually assumed as American English) produced by Caucasian speakers. There are few cross-language studies on elderly speech. It is also rare that linguistic theories are well-incorporated in these studies. This chapter reviews the findings and shortcomings of research on age-related speech characteristics and the perception of vocal age.

Age perception research

There is no doubt that elderly speech is different from speech produced by young people. The current consensus of age perception research is that listeners can tell quite easily whether the speaker is young or old just by listening to their speech. A number of studies confirm that listeners’ estimation of speaker’s age is indeed quite accurate (e.g. Ptacek & Sander, 1966; Shipp, 1973; Ryan & Burk, 1974; Neiman & Applegate, 1990; Linville & Fisher, 1985a; Huntley, Hollien, & Shipp, 1987; Horii & Ryan, 1981; Hartman, 1979; Capadano, 1978; Braun & Rietveld, 1995; Jacques & Rastatter, 1990).

The first experimental study was conducted by Ptacek and Sander in 1966. They asked ten listeners to categorize the speakers’ age as either young (under 35) or old (over 65) in three conditions: sustained vowel /a/, backward-play speech, and forward-play speech. Samples of a sustained vowel were collected from 72 speakers and the speech samples were from 36 speakers. Age groups and speakers’ gender were balanced in each set of speech samples. Their young speakers were 18 to 34 years old, and the old speakers were 67 to 87 years old. Speakers were asked to produce the vowel at the same pitch level. The frequencies employed for male and female speakers were 130 (±30) Hz and 210 (±30) Hz respectively. Intensity level was also monitored to be around 82 dB using visual feedback of a VU meter. The average correct identification scores for
sustained vowels, backward speech, and forward speech were 78, 87, and 99%, respectively. The result that the listeners can categorize speakers’ age from prolonged vowels alone suggests that laryngeal characteristics provide perceptual cues associated with elderly speech. This was supported later by the results that whispered vowels did not provide enough perceptual cues of speaker age to listeners (Linville & Fisher, 1985a; Jacques & Rastatter, 1990). Increased accuracy in forward speech suggested other speech characteristics also played an important role in age perception. In the backward speech condition, listeners tended to misidentify young speech as old (72% of the total errors), but not the other way around. Errors could be due to the oddity of the stimulus. However, the interesting point was that when the listeners misidentified the speaker’s age, they assigned it to the old category. The uni-directional error pattern (misidentification of young speech as old) could imply that whenever the speech sounded odd, the listeners tend to categorize it as old. Hence, elderly speech and some odd quality may be more likely associated with each other.

In 1969, Shipp and Hollien conducted a similar study. In addition to categorizing speaker into age groups, a direct age estimation task was introduced. Shipp and Hollien collected speech samples from 25 male speakers in each decade from 20 to 89 years. Their study is still the largest study on age perception in terms of the sample size and the age range of the speakers. Three groups of twenty listeners were employed for three different age estimation tasks. One group of listeners was asked to estimate speakers’ age using the 3-point scale representing from “young” (=1), “neither old nor young” (=2), and “old” (=3). Another group was assigned to estimate speakers’ age with the 7-point scale representing decade of age range from 20 to 80 years. The last group of listeners
was asked to estimate speakers’ age directly. Their results replicated the finding of Ptacek and Sander (1966) that the listeners estimate speaker’s age accurately. Correlation between chronological age and average perceived age in the direct age estimation was $r = 0.88$. They also observed that listeners tend to underestimate the age of older speakers and overestimate the age of younger speakers.

Several studies observed the same phenomenon that listeners tend to underestimate the age of older speakers (Hollien & Tolhurst, 1978; Shipp & Hollien, 1969; Braun, 1996; Hartman, 1979) and overestimate the age of younger speakers (Shipp & Hollien, 1969; Huntley et al., 1987; Braun, 1996). One of the reasons for these misjudgments could be due to the usage of a narrower age scale for perceived age than chronological age. The judges used the extreme ends of age decade for only a few talkers. Further analysis of inter-judge agreement revealed that the listeners had higher agreement on their judgments for younger speakers than older speakers. Considering that the younger speakers could be a peer group of listeners, Shipp and Hollien suggested that familiarity with individual voices could affect age perception. Overall results of previous studies suggest that listeners can identify speakers’ age and that there are certain perceptually identifiable parameters that are associated with age in speech.

**Age-related acoustic characteristics of voice and speech**

Compared to perceptual studies of vocal age, more comprehensive investigations have been done for acoustic characteristics of speech and voice associated with advanced aging. Previous literature suggests that there are several acoustic characteristics of elderly voice and their speech sounds. The causes of these differences are attributed to age-related sensorimotor changes or physiological deterioration seen in vocal apparatus.
Fundamental frequency

Among the many acoustic variables, the most extensive research has been done on average fundamental frequency. Voluminous data on fundamental frequency has supported that mean speaking fundamental frequency continues to change after puberty. Although results are not unanimous, the majority of investigations appear to support age-related changes in fundamental frequency that are different between men and women. In general, fundamental frequency in male speech decreases after puberty, remains in the same range or slightly decreases during mid-adulthood, and increases with advancing aging (Mysak, 1959; Mysak & Hanley, 1958; Hollien & Shipp, 1972; Xue & Deliyski, 2001; W. S. Brown, Morris, Hollien, & Howell, 1991; Harnsberger, Shrivastav, Brown, Rothman, & Hollien, In press). There is a disagreement among researchers for females. The majority of research suggests that mean fundamental frequency slightly decreases from early adulthood (Fitch & Holbrook, 1970) to middle-age (Stoicheff, 1981; Saxman & Burk, 1967) to advanced age (Krook, 1988; Xue, Neeley, Hagstrom, & Hao, 2001; Xue & Mueller, 1996; Xue & Deliyski, 2001; Honjo & Isshiki, 1980; de Pinto & Hollien, 1982; W. S. Brown et al., 1991; Yamazawa & Hollien, 1992). However, McGlone & Hollien (1963), Gilbert & Weismer (1974), and Coats-Garret (1980) observed little change in mean fundamental frequency as a function of advancing age. It should be noted that smoking affects fundamental frequency. At least among middle-aged females, women with a smoking habit have lower fundamental frequency than the non-smoking women (Gilbert & Weismer, 1974).

Most of the investigations on mean fundamental frequency have focused on Caucasian subjects. Speakers with different race or speakers of a non-English native
language are known to exhibit different mean fundamental frequencies from ones observed for Caucasian English speakers. A few studies reported lower fundamental frequency for young African-American speakers compared with age-matched Caucasian speakers (Hudson & Holbrook, 1981, 1982; Hollien & Maleik, 1962; Ducote, 1983). This difference was found in the elderly female speakers, but not for the elderly male speakers (Xue & Mueller, 1996; Xue et al., 2001). Studies on Japanese speakers support a general trend of age-related fundamental frequency changes, that is a gradual decrease for female speakers (Yamazawa & Hollien, 1992; Honjo & Isshiki, 1980; Tsuge, Kakami, & Fukaya, 1987; Terasawa, Kakita, & Hirano, 1984; Hanley & Snidecor, 1967) and an increase for male speakers, as expected from previous research on English speakers. However it should be mentioned that Japanese female speakers tend to exhibit higher average speaking fundamental frequencies than English female speakers (Yamazawa & Hollien, 1992; Loveday, 1981). It is also known that Japanese female speakers employ higher pitch range when they speak in Japanese than when they speak in English. But there are no significant differences in the speaking fundamental frequency of Japanese male productions in English and Japanese (Ohara, 1992). This gender difference in pitch is attributed to sharper distinctions on gender role in Japanese society (but see Yamazawa & Hollien, 1992).

Large individual variability is a general characteristic observed in elderly populations. Intrasubject variability and intersubject variability of mean fundamental frequency are large among the elderly speakers. A typical measure for variability is standard deviation of mean fundamental frequency ($F_0$SD) during prolonged vowel phonation. Many investigators observed large $F_0$SD among elderly speakers and within
an elderly speaker (B. J. Benjamin, 1981). Intersubject variability can be reduced when
the subjects were grouped by physiological age, not chronological age (Ramig, 1983a,
1983b; Ramig & Ringel, 1983).

F0SD during sustained vowel phonation was also used as a measure for degree of
shakiness or presence of voice tremor. Elderly voice is often described as being more
breathy, harsh, rough, or shaky. Several acoustic correlates of these voice qualities have
been extensively examined. Some investigators used shimmer and jitter as measures for
breathiness, harshness, or roughness, and some have used harmonics-to-noise ratio
(HNR) as a measure for hoarseness quality of voice (Braun & Rietveld, 1995). Ramig &
Ringel (1983) conducted acoustic analysis of elderly speech to examine whether there is
any difference in the irregularities in pitch and amplitude (shimmer and jitter). They
found that the average values of these two measures were greater in the elderly speakers
than the younger speakers. More recently, Xue and Deliyski (2001) reported that elderly
speakers produce a sustained vowel with more perturbation in their fundamental
frequency than younger speakers.

Temporal characteristics

Temporal characteristics of speech have also been often mentioned with aging
speech. The motivation behind this is slowed sensorimotor skills among elderly people.
These include speech rate, syllable rate, segmental duration, and frequency of pause.
Researchers agree that speech by the elderly speakers has a slow rate. In terms of words
spoken per minute in reading, older speakers read the text at slower rate than the middle-
age speakers (Mysak, 1959; Ryan, 1972; B. L. Smith, Wasowicz, & Preston, 1987;
Ramig, 1983a). Elderly speakers tend to produce fewer syllables per breath group
Segment durations become longer with age (B. J. Benjamin, 1982; Ramig, 1983a; Weismer & Fromm, 1983; Harnsberger et al., In press). However, elderly speakers have shorter voice onset time for voiceless stops (B. J. Benjamin, 1982; Neiman, Klich, & Shuey, 1983). As for the voiced counterpart, elderly speakers tend to produce English voiced stops with voicing lead more often than younger speakers (Morris & Brown, 1987). Shorter VOT for stops is considered as a result of deteriorated coordination of glottal and supraglottal timing (Weismer & Fromm, 1983; Weismer, 1984). Although Sweeting and Baken (1982) did not find VOT differences across the three age groups they compared, they observed more variability in elderly VOT values than what is seen in younger productions (Sweeting & Baken, 1982). Furthermore, the opposite phenomenon was reported in Japanese (Shimizu, 1996; Homma, 1980; Shimizu, 1999; Sugito, 1996). It has been documented that Japanese /d/ has negative VOT. However, recent study indicated that younger Japanese talkers tend to produce /d/ with more positive VOT than older talkers (Takada, 2004). Takada (2004) proposed that the age-related difference in occurrence rate of voiceless /d/ is a part of active historical sound change.

Resonant characteristics

Compared to the age differences observed in pitch, resonant characteristics do not change much with aging, although slight differences in vowel formant frequency have been observed between young and elderly speakers (Xue & Hao, 2003; Linville & Fisher, 1985b). Linville and Fisher (1985b) observed that elderly female speakers exhibited lower F1 values than young female speakers during the phonated vowel /æ/. For elderly male speakers, however, both F1 and F2 values did not deviate much from the typical
values found for younger adults in other literature (Liss, Weismer, & Rosenbek, 1990). More recently, Xue and Hao (2003) investigated the age effect on formant frequencies in vowels produced by men and women, and found that elderly talkers overall exhibited lower formant frequencies than young talkers.

The actual cause of formant lowering has not been determined yet because of inconsistency among research findings, but physiologically based accounts have been given so far. Linville and Fisher (1985b) suggested that they are due to increased size of the vocal tract due to age-related changes in the larynx. However, Rastatter and Jacques (1990) reported conflicting results with Linville & Fisher (1985b) and Xue & Hao (2003). Rastatter and Jacques (1990) investigated differences in F1 and F2 formant frequency values in vowel productions of young and elderly male and female speakers, and did not find systematic formant lowering in elderly talkers. Instead, they found that elderly talkers exhibited a reduced vowel space relative to the vowel speech in young talkers. They suggested that the age-related changes in tongue positions can account for differences in formant values between young and elderly talkers. It is difficult to compare these studies due to the methodological differences. Vowels were produced in an isolated context in Linville and Fisher (1985b), Rastatter & Jacques (1990), and Xue & Hao (2003), but in a sentence in Liss et al. (1990). Although Rastatter, McGuire, Kalinowski, and Stuart (1997) found no significant effects of neighboring consonants, this contextual differences might be the reason of differences in these studies.

In addition, these studies did not fully discuss a linguistic factor to account for the formant differences between young and elderly speakers although Linville and Fisher (1985b) briefly mentioned a possible effect from dialect variation. However, it has been
documented that vowels undergo active sound changes, e.g. the Northern Cities Vowel Shift in American English (e.g. Labov, Yaeger, & Steiner, 1973; Labov, 1994). For example, diachronic changes can explain portions of results in Rastatter et al. (1997). They found that elderly male talkers exhibited higher F1 for the vowel /u/ and higher F2 for the vowel /æ/ than young male talkers, which was opposite the direction of change due to aging, if formant frequency changes as a function of the size of the vocal tract that lengthened with advancing aging. Ongoing sound changes such as raising of /æ/ and fronting of /u/ in American English (Labov, Ash, & Boberg, 2005) could account for these results because it predicts that younger talkers would pronounce the /æ/ vowel in more front and higher position and the /u/ vowel in more front position than elderly talkers. Hence, conflicting results might be due to regional or social variations in vowels.

**Linguistic characteristics**

In spite of various purposes and implications of the studies cited above, little is known about elderly speech from a linguistic point of view. Many acoustic studies have investigated the voice characteristics of elderly people. Although the upper limits of human ability in speech production, such as duration of sustaining phonation or rapid movements of articulators, decreases with advancing age, speech does generally not require these extreme capabilities. Detailed acoustic examination of elderly speech has rarely included the suprasegmental level. Thus it is not known whether or not speech characteristics of elderly speakers might be problematic for conveying linguistic information.
Previous research proposed physiological changes associated with advanced aging based on correlations computed between perceptual results and acoustic characteristics of aged speech, but there are limitations to how far one can generalize these results to all aspects of speech by elderly. It seems that researchers tend to underestimate the communicative abilities of elderly people with normal aging. One of the reasons relates to the terminology used for perceptual attributes. The labels used for descriptions of voice characteristics were originally used to diagnose communicative disorders (Darly, Aronson, & Brown, 1969). Especially, after Ryan and Burk (1974, p.191) remarked that “individuals judged to be older may be speakers who should fall at the mild end of a “dysarthria continuum”, investigators are sometimes prejudiced in favor of treating elderly speech as something deviant from the normal speech. However, speech by the elderly speakers does not differ from speech produced by younger speakers for all aspects of speech. For example, overall fluency does not decrease with aging. Leeper and Culatta (1995) computed the total disfluency scores for the reading of the “Rainbow Passage” by four groups of old speakers and one control group of the young speakers. They found that there were no differences even between the young and the oldest group of speakers in terms of disfluency. Slightly increased disfluency in reading compared to spontaneous speech could be the result from reduced visual acuity in the elderly. Even centenarians exhibit no significant differences in terms of speech disfluency (Caruso, McClowry, & Max, 1997; Searl, Gabel, & Fulks, 2002).

However, a decline of sentence complexity was observed in elderly sentence productions. Simple sentence constructions included less use of relative clauses, that-
clauses, wh-clauses, double or triple embeddings, and infinitive complements (Kemper, 1988).

Perceptual and acoustic correlates of age perception

The question of what kind of perceptual and acoustic cues enabled listeners to perceive speakers’ age was explored by Ryan and Burk (1974). They collected speech samples from 80 male speakers whose ages range from 40 to 80 years old. Twenty female listeners were able to estimate the age of speakers with high accuracy similar to the previous studies ($r = 0.8$ to $0.95$). Then, eighteen trained listeners judged whether certain voice characteristics were present or absent. Ten voice characteristics used in this study were air loss, laryngeal tension, vocal fry, pitch breaks (sudden and uncontrolled variation in pitch), voice tremor (shakiness or tremulousness in voice), hypernasality, hyponasality, imprecise consonants, slow rate, and slow rate of articulation. It should be noted that most of the voice characteristics in Ryan and Burk’s study had been employed to describe disordered speech. Five acoustic characteristics (mean dB SPL, words per minute, words per minute per sentence, mean fundamental frequency, and SD of fundamental frequency) were also measured. The most highly correlated variable with aging speech was voice tremor. Among the five acoustic characteristics examined in this study, standard deviation of fundamental frequency was the only variable primarily related to voice tremor. However, variability in fundamental frequency was not correlated with perceived age. Mean fundamental frequency also did not correlate well with age perception. Similarly, Braun and Rieveld (1995) reported that the speaking fundamental frequency (SFF) was not a good predictor of perceived age. More recently, Harnsberger et al. (In press) conducted the age estimation experiments using
resynthesized speech stimuli, where fundamental frequency and/or speaking rate were either increased or decreased from the original speech produced by young, middle-aged, and elderly men. They found that the manipulation of speaking rate affected the perceived age for the young talkers as well as the middle-aged and elderly talkers, but the effects of fundamental frequency manipulation were only found for the elderly talkers. This result was not expected from acoustic analysis of aging speech because speaking fundamental frequency changes with age in general. In fact, in contrast to the Ryan and Burk’s finding, Horii and Ryan (1981) found positive correlations between mean fundamental frequency and perceived age for male speakers with an age range from 40 to 80 years. Standard deviation of fundamental frequency was not found to be correlated with perceived age in their study.

Lack of association between mean fundamental frequency and perceived age might be due to methodological inadequacy. Reading of the entire text of the “Rainbow Passage” from each speaker was used in the perceptual judgment experiment. Since it was from lengthy material that means and standard deviations of fundamental frequency were measured, both measures could vary within the individual words or sentences. When mean or standard deviation of fundamental frequency was taken as an indirect measure for perceptual cues such as voice tremor, it is not surprising that they failed to show correlations between perceptual cues and acoustic measures. Acoustical features that are expected to be correlated with perceptual characteristics of voice quality should be taken from more homogeneous speech samples such as prolonged phonation.

Difference in voice quality (e.g. breathy, shaky, strained, etc.) is one of the most salient perceptual cues for aging speech (Ryan & Burk, 1974; Kukol, 1979). Although
only a few studies have investigated the acoustic correlates of voice quality and perceived age, several studies support the relationship between perceived age and acoustic measures is related with voice qualities. Significant correlation was found between perceived age for female talkers and F0 standard deviation (acoustic correlates of vocal tremor or ‘wobbliness’) during the vowel phonation (Linville & Fisher, 1985a). Braun and Rietveld (1995) reported that harmonic-to-noise ratio (acoustic correlate of roughness) was a good predictor of perceived age. Spectral tilt is related with creaky or breathy voice quality. Kido and Kasuya (Kido & Kasuya, 2005) found significant correlation between spectral tilt and perceived age for Japanese talkers. However, long-term averaged spectra (another acoustic correlate for breathiness) were not significantly correlated with perceived age (Linville, 2002). Research varies in terms of the stimuli types and it is difficult to compare those results. Further investigations will be needed in order to understand the perceptual and acoustic correlates of vocal age.

Similarly, there is a sparse literature on the relationship between perceived age and resonant characteristics. From the comparative study of perceived age for whispered vowels and phonated vowels, results showed that the resonant characteristics (F1 and F2) do not contribute to perception of speaker age (Linville & Korabec, 1986; Linville & Fisher, 1985a).

As for temporal characteristics, either a weak or a strong correlation was found between perceived age and speech rate (operationally defined as words or syllables per minute; Braun & Rietveld, 1995; Ryan & Burk, 1974; Ptacek & Sander, 1966). Listener’s perceptual judgment on slowness of speech is a good predictor of perceived age. The relationships between acoustic measures for speech rate and perceived age are
correlated well in English (Shipp, Qui, Huntley, & Hollien, 1992), German (Braun & Rietveld, 1995), and Japanese (Kido & Kasuya, 2005). Interestingly, Ryan and Burk (1974) did not find a significant correlation for the acoustic variable of speaking rate (words per minute) although the perceptual variable for ‘slow rate of articulation’ was strongly correlated with perceived age. However, results in the previous studies seem to suggest that speaking rate influences listener’s judgments on a speaker’s age. Recent perception study with resynthesized stimuli also supports that the speaking rate is a strong perceptual cue to identify the age in speech.

**Other factors influencing age perception**

Other than acoustic factors, previous literature has shown that there are other factors to influence age perception such as listener age, listener sex, smoking effects, and other experiment procedural factors, as will be seen in detail below.

**Speech materials**

Different amounts of linguistic information in each stimulus seem to influence perception of the talker’s age. As far as we know, there are only two studies that used more than two types of stimuli that included different degrees of linguistic information (Ptacek & Sander, 1966; Neiman & Applegate, 1990). Ptaceck and Sanders (1969) used sentence stimuli and isolated vowel stimuli, and found an increased accuracy of age categorization for sentences over vowels. However, the task employed in their study was not a direct age estimation task but a simple age categorization task, where the listeners differentiated whether the speaker was young or old. Although the listeners identified the
age group of talkers more accurately when they listened to the sentence stimuli than the vowel stimuli in their study, it was not clear how finely listeners differentiated the talker’s age in each type of stimuli. Neiman and Applegate (1990) included two types of stimuli, phrases and sentences, in their age perception study. Their results seem to suggest that age estimation was more accurate for sentences than phrases. However, the results for each type of stimuli were available only from the reliability test, and they did not further analyze the results in terms of the types of speech materials; hence, it is difficult to determine whether the listeners estimated the age of talkers more accurately for sentences than phrases.

Table 1 lists previous studies on age perception and their results, grouped by the stimuli type they employed. Most of the previous studies employed the stimuli from the reading of sentences, while some studies (Ramig, Scherer, & Titze, 1985; Linville & Fisher, 1985a; Linville & Korabic, 1986) used vowel stimuli instead of sentence stimuli. An interesting point is that the worst and only non-significant correlation result was found for the vowel stimuli, i.e., $r=0.17$ in Ramig et al. (1985). On the contrary, in the studies that used the sentence stimuli, correlations between chronological age and perceived age were high, ranging from 0.66 to 0.96. Although correlation between chronological age and perceived age was low with the vowel stimuli in Ramig et al. (1985), listeners were able to identify the age groups of talkers by listening to sustained vowels. Yet, the sentence stimuli clearly provided the listeners more information about a talker’s age. These results suggest that age perception is not based on voice quality alone because the vowel stimuli obviously did not include sufficient information to estimate the age of talkers.
Improved age estimation from the vowel to the sentence stimuli suggests that information available in the sentence stimuli were effective perceptual cues for age estimation. Important information in the sentence stimuli would be temporal information such as pauses and speaking rate and some linguistic information such as phrasing or intonation. Extremely good age estimation for conversational speech in Hartmann (1979) may be related to individual differences in lexical choices since diachronic changes in vocabulary and language use provide valuable age information to listeners. However, the listeners could estimate a talker’s age without any differences in lexical items in the stimuli. It is natural to conclude that listeners’ age estimation became more accurate when they were given more speech information to evaluate. Then, the question is how much and what kinds of information are necessary for listeners to be able to estimate the age of talker at a certain level of accuracy.
Table 1. Types of the stimuli used in previous studies on age perception. Correlation values were mean values except Kukol (1979), who reported median correlation values. Accuracy was measured in terms of correct percent of age group identifications (with chance probability in the parenthesis) in some studies, or by the average age differences between chronological age and perceived age in other studies. Correlations were not available in the majority of studies employed categorization tasks of two or three age groups. Accuracy results were also not reported in some of the studies.

<table>
<thead>
<tr>
<th>Source</th>
<th>Stimuli type</th>
<th>Correlations</th>
<th>Accuracy (% chance level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ptaceck &amp; Sanders (1966)</td>
<td>Vowel</td>
<td>n/a</td>
<td>78% (50%)</td>
</tr>
<tr>
<td>Linville &amp; Fisher (1985)</td>
<td>Vowel</td>
<td>n/a</td>
<td>51% (33%)</td>
</tr>
<tr>
<td>Linville &amp; Korabic (1986)</td>
<td>Vowel</td>
<td>n/a</td>
<td>45% (33%)</td>
</tr>
<tr>
<td>Jacques &amp; Rastatter (1990)</td>
<td>Vowel</td>
<td>0.17</td>
<td>n/a</td>
</tr>
<tr>
<td>Ramig et al. (1985)</td>
<td>Vowel</td>
<td>0.17</td>
<td>n/a</td>
</tr>
<tr>
<td>Neiman &amp; Applegate (1990)</td>
<td>Phrase</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Ptaceck &amp; Sanders (1966)</td>
<td>Sentence</td>
<td>n/a</td>
<td>99% (50%)</td>
</tr>
<tr>
<td>Cerrato et al. (2000)</td>
<td>Sentence</td>
<td>0.77</td>
<td>84% (14%)</td>
</tr>
<tr>
<td>Shipp &amp; Hollien (1969)</td>
<td>Sentence</td>
<td>0.88</td>
<td>n/a</td>
</tr>
<tr>
<td>Ryan &amp; Burk (1974)</td>
<td>Sentence</td>
<td>0.77¹</td>
<td>n/a</td>
</tr>
<tr>
<td>Horii &amp; Ryan (1981)</td>
<td>Sentence</td>
<td>0.76</td>
<td>n/a</td>
</tr>
<tr>
<td>Baker (1981)</td>
<td>Sentence</td>
<td>0.68¹(Caucasian listeners)</td>
<td>−12.1 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.69¹(Afr.-Am. listeners)</td>
<td>−13.2 years</td>
</tr>
<tr>
<td>Kukol (1979)</td>
<td>Sentence</td>
<td>0.80* (male speakers)</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.63* (female speakers)</td>
<td>n/a</td>
</tr>
<tr>
<td>Neiman &amp; Applegate (1990)</td>
<td>Sentence</td>
<td>0.88 (male speakers)</td>
<td>−10.2 years²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.96 (female speakers)</td>
<td>−8.6 years²</td>
</tr>
<tr>
<td>Braun (1996)</td>
<td>Sentence</td>
<td>0.77 (untrained listeners)</td>
<td>−6.5 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.70 (trained listeners)</td>
<td>−5.9 years</td>
</tr>
<tr>
<td>Kido &amp; Kasuya (Kido &amp; Kasuya, 2004)</td>
<td>Sentence</td>
<td>0.66</td>
<td>[7.2] years</td>
</tr>
<tr>
<td>Hartmann (1979)</td>
<td>Conv. Sp.</td>
<td>0.94</td>
<td>−5.8 years³</td>
</tr>
<tr>
<td>Mulac &amp; Giles (1996)</td>
<td>Conv. Sp.</td>
<td>0.51</td>
<td>−8.3 years</td>
</tr>
</tbody>
</table>

¹ Correlation values were calculated based on the reported individual data by the author.
² Correspondence between perceived age and chronological age was converted to the age difference between perceived age and chronological age in years by the author.
³ Differences between chronological age and perceived age (=PA−CA) were calculated by the author.

[Conv. Sp. = conversational speech; Afr.-Am.=African-American]
Sentence stimuli include highly robust speech characteristics related to aging, while vowel stimuli provide voice characteristics mainly determined by physiological changes associated with aging. The reason that some acoustic correlates of age perception are not conclusive among researchers may be rooted in normal speech variability among individual sentences. Especially, prosodic characteristics vary among talkers. Even if the same text was used for all talkers, each talker could organize prosodic structures in a different way. Specifically, the phrasing of utterances would significantly affect prosodic characteristics among individual talkers. For example, when a sentence was read with many pauses instead of reading without a pause, the average fundamental frequency may increase due to multiple phrases in the utterance, because declination of pitch that is naturally observed in a declarative type of sentence will be reset at the beginning of every new phrase. Also, individual differences at the segmental level will be more apparent in the sentence stimuli. Reduced articulation of vowels or consonants will occur more likely in the sentence materials. In addition, socio-phonetic variations can be a good source of age information in many dialects. Pronunciation of some vowels changes over time, e.g., the Northern Cities Vowel Shift in American English (e.g. Labov et al., 1973; Labov, 1994).

As a result of such historical change, different age groups in the same language community pronounce the same phoneme differently. Such socio-phonetic variations could indicate a talker’s age if both talkers and listeners are from the same dialect region. A talker’s age is an important variable to explain occurrences of phonological variations as well (e.g. Labov, 1972 for vowel centralization; e.g. Guy, 1980; Tagliamonte & Temple, 2005 for /t/-/d/ deletion). However, because the detailed dialect information of
talkers and listeners were not described in previous studies on age perception, the effect of sociolinguistic variations on age estimation is not clear.

**Speakers**

**Speaker’s native language**

As it happens in many fields of study, age perception literature is dominated by studies of American English. Only a handful of studies have been done for languages other than English. Brawn (1996) conducted an age estimation experiment for a reading of the German version of *the North Wind and the Sun* by forty males whose age ranged from 25-59 years. High correlations between speakers’ chronological age and perceived age were found for twelve trained listeners (mean age =40.7) and nineteen untrained listeners (mean age =23.3), $r= 0.70$, $r= 0.68$, respectively. Their correlations were not as high as seen in previous studies. Exclusion of elderly speakers might be one of the reasons for this result. Cerrato, Falcone, and Paoloni (2000) investigated listeners’ ability for age estimation of telephone speech for forensic purposes. They used speech samples collected from 42 Italian speakers over telephone. Three female and three male speakers were selected from seven age groups from 18 to 66 years. Seventeen native listeners of Italian estimated speakers’ age by selecting one of the seven age groups they employed. Listeners’ perceived age and speakers’ chronological age correlated well ($r= 0.70$). Kido and Kasuya (2004) examined relationship between perceived age and chronological age in Japanese, and also found a moderate correlation between the two ($r= 0.66$).

Across these languages, results were consistent that listeners were able to estimate the age of talkers accurately. **Table 2** lists correlations between chronological age and
perceived age in these studies Correlation values in those three non-English languages were high, and similar to correlation values observed in English.

<table>
<thead>
<tr>
<th>Source</th>
<th>Language</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shipp &amp; Hollien (1969)</td>
<td>AE</td>
<td>0.88</td>
</tr>
<tr>
<td>Ryan &amp; Burk (1974)</td>
<td>AE</td>
<td>0.77*</td>
</tr>
<tr>
<td>Horii &amp; Ryan (1981)</td>
<td>AE</td>
<td>0.76</td>
</tr>
<tr>
<td>Baker (1981)</td>
<td>AE</td>
<td>0.68* (Caucasian listeners)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.69 (African-American listeners)</td>
</tr>
<tr>
<td>Kukol (1979)</td>
<td>AE</td>
<td>0.80 (male speakers)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.63 (female speakers)</td>
</tr>
<tr>
<td>Neiman &amp; Applegate (1990)</td>
<td>AE</td>
<td>0.88 (male speakers)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.96 (female speakers)</td>
</tr>
<tr>
<td>Braun (1996)</td>
<td>German</td>
<td>0.77 (untrained listeners)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.70 (trained listeners)</td>
</tr>
<tr>
<td>Cerrato et al. (2000)</td>
<td>Italian</td>
<td>0.77</td>
</tr>
<tr>
<td>Kido &amp; Kasuya (2004)</td>
<td>Japanese</td>
<td>0.66</td>
</tr>
<tr>
<td>Hartmann (1979)</td>
<td>AE</td>
<td>0.94</td>
</tr>
<tr>
<td>Mulac &amp; Giles (1996)</td>
<td>AE</td>
<td>0.51</td>
</tr>
</tbody>
</table>

* Correlation values were calculated based on the reported individual data by the author.

Age estimation errors were also similar across languages. Remember that English listeners tend to underestimate age of older talkers and overestimate age of younger talkers. Cerrato, Falcone, and Paoloni (2000) reported that “listeners tend to overestimate young speakers and to underestimate old speakers”. Japanese also shows the same types of age estimation errors in Kido & Kasuya (2004).

As far as the author knows, age perception research has not been conducted across languages. G. R. Benjamin (1992) may be the only cross-language comparison study of age perception. She investigated perception of age in English and Japanese using spontaneous speech, but the effect of speaker’s native language was not directly
examined because the listeners estimated the age of talkers in their native language. It is also difficult to evaluate age estimation accuracy in either language group, because the information about the age of talkers was not collected in her study.

Comparison of age perception, however, was conducted across different ethnicities by Baker (1981). This appears to be the only known perception study involving race, specifically a study on African-American speech. Baker (1981) conducted a direct age estimation experiment with speech (The Rainbow Passage) produced by 82 African-American male speakers and 90 African-American female speakers. Female speakers were from 40-95 years old, and male speakers were 40-81. Twenty African-American female listeners and twenty Caucasian female listeners were instructed to do direct age estimation of the speakers. Overall results support previous studies of age perception. Two listener groups showed no differences in perceived age for female speakers. However, Caucasian listeners estimated the age of African-American male speakers as younger than estimates of the African-American listeners. This suggested that ethnicity may play a role in perception of speaker age. If the differences in speech due to ethnicity influenced the listener’s perception, it is quite possible that language affects on age perception.

**Speaker age**

Table 3 shows the age range of talkers in various studies on age perception. Range and mean of age employed as targeted group of speakers varies among studies, even though age of speakers is an important factor. The most common demarcation of age between old and non-old speakers is 65 years old. But some investigators refer to people in their fifties as older people. Similarly, people at age of 35 can be grouped as
either a young group or a middle-aged group. In this table, the age range of target
speakers ranges from four decades to seven decades. In some studies (e.g. Shipp &
Hollien, 1969), talkers were carefully selected to collect speech samples at each age (in
years), while only limited number of age groups was selected in other studies.

The age range of talkers potentially affects correlation results. Because listeners
were given the expected age range of talkers, the range of age estimation was bounded
within the age range of talkers. When the restricted range is used for responses, the
correlation value is lower than it would be if the complete range were sampled (McCall,
1998). Correlations between chronological age and perceived age were moderate to high
in some of the previous studies (Ryan & Burk, 1974; Horii & Ryan, 1981; Mulac & Giles,
1996) even though the age range of talkers was relatively narrow (e.g. 40-80 years old or
59-92 years old). However, similarly a high correlation was found when the listeners did
not know the expected age range (Baker, 1981).

<table>
<thead>
<tr>
<th>Source</th>
<th>Correlations between PA and CA</th>
<th>Age range of speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shipp &amp; Hollien (1969)</td>
<td>0.88</td>
<td>20-89</td>
</tr>
<tr>
<td>Ryan &amp; Burk (1974)</td>
<td>0.77*</td>
<td>40-80</td>
</tr>
<tr>
<td>Horii &amp; Ryan (1981)</td>
<td>0.76</td>
<td>40-80</td>
</tr>
<tr>
<td>Baker (1981)</td>
<td>0.68* (Caucasian listeners)</td>
<td>40-95</td>
</tr>
<tr>
<td>Neiman &amp; Applegate (1990)</td>
<td>0.91</td>
<td>20-75</td>
</tr>
<tr>
<td>Braun (1996)</td>
<td>0.68 (untrained listeners)</td>
<td>25-59</td>
</tr>
<tr>
<td></td>
<td>0.70 (trained listeners)</td>
<td></td>
</tr>
<tr>
<td>Ramig et al. (1985)</td>
<td>0.17</td>
<td>25-75</td>
</tr>
<tr>
<td>Neiman &amp; Applegate (1990)</td>
<td>0.91</td>
<td>20-75</td>
</tr>
<tr>
<td>Braun (1996)</td>
<td>0.68 (untrained listeners)</td>
<td>25-59</td>
</tr>
<tr>
<td></td>
<td>0.70 (trained listeners)</td>
<td></td>
</tr>
<tr>
<td>Cerrato et al. (2000)</td>
<td>0.77</td>
<td>18-66</td>
</tr>
<tr>
<td>Neiman &amp; Applegate (1990)</td>
<td>0.91</td>
<td>20-75</td>
</tr>
<tr>
<td>Braun (1996)</td>
<td>0.68 (untrained listeners)</td>
<td>25-59</td>
</tr>
<tr>
<td></td>
<td>0.70 (trained listeners)</td>
<td></td>
</tr>
<tr>
<td>Cerrato et al. (2000)</td>
<td>0.77</td>
<td>18-66</td>
</tr>
<tr>
<td>Mulac &amp; Giles (1996)</td>
<td>0.51</td>
<td>59-92</td>
</tr>
<tr>
<td>Kido &amp; Kasuya (2004)</td>
<td>0.66</td>
<td>20-60</td>
</tr>
</tbody>
</table>

* Correlation values were calculated based on the reported individual data by the author.
Speaker sex

It is common that research on elderly speech has drawn general conclusions based on the speech sampled from only one gender. However, as mentioned above, acoustic characteristics associated with advancing aging are different between males and females. Baker (1981b) and Neiman and Applegate (1990) found that listeners estimate age more accurately when they listen to female speech than male speech. However, the correlation value for female speech in Kukol (1979) was lower than the correlation values for female talkers in Baker (1981b) and Neiman and Applegate (1990). Moreover, correlations for male speech (Shipp & Hollien, 1969; Ryan & Burk, 1974; Hartmann, 1979) were similar or higher than correlations observed for female speech (Baker, 1981; Neiman & Applegate, 1991; Kukol, 1979), as seen in Table 4. It is not determined from comparing previous studies whether or not speaker sex affects listeners’ age estimation.

<table>
<thead>
<tr>
<th>Source</th>
<th>Correlation for male speakers</th>
<th>Correlation for female speakers</th>
<th>Listener sex (Race/Ethnicity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shipp &amp; Hollien (1969)</td>
<td>0.85</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Ryan &amp; Bark (1974)</td>
<td>0.90</td>
<td>F (Caucasian)</td>
<td></td>
</tr>
<tr>
<td>Burk, Hoyer &amp; Fey (1975)*</td>
<td>0.80</td>
<td>M and F (Caucasian)</td>
<td></td>
</tr>
<tr>
<td>Hartmann (1979)</td>
<td>0.93-0.94</td>
<td>M and F</td>
<td></td>
</tr>
<tr>
<td>Kukol (1979)</td>
<td>0.66**</td>
<td>F (Caucasian)</td>
<td></td>
</tr>
<tr>
<td>Baker (1981)</td>
<td>0.68</td>
<td>0.82</td>
<td>F (Caucasian)</td>
</tr>
<tr>
<td>Neiman &amp; Applegate (1990)</td>
<td>0.88</td>
<td>0.96</td>
<td>F (African-American)</td>
</tr>
</tbody>
</table>

* Data was taken from Table 25 in Baker (1981).
** Median r.
Listeners

Listener age

As for listener age, almost exclusively young adults (20-35 years) were recruited for the perception experiments. However, listener age has been shown to influence age perception. Huntley, Hollien and Shipp (1987) examined the effect of listener age on perceived age. Listeners were recruited from four age groups for the direct age estimation of 105 male speakers with age range of 20-90 years. Age ranges of four listener groups were: 9-15 years in the adolescent group, 20-30 years in the young listener group, 40-50 years in the middle-aged listener group, and 60-84 years in the old listener group. Listeners’ performance of age estimation was as good as in previous research. They also found similar error judgments, that is, listeners tended to underestimate the age of older speakers, whereas they tended to overestimate the age of younger speakers. No significant difference was found among the listener groups in terms of age estimation accuracy, except age estimation for the younger speakers. Specifically, overestimation for the younger speakers was more evident in the adolescent and old listener groups than the young and middle-aged listener groups. Huntley et al. (1987) suggested that a possible explanation of poor performance by adolescents to estimate the age of young speakers is their underdeveloped conceptualization of age. One of the possible explanations for older listeners’ poor performance for younger speakers was their lack of social contact with younger people.

On the contrary, Jacques and Rastatter (1990) found a different pattern of results. They investigated the effects of age, sex, and vowels on speaker age recognition. Four groups of listeners (young male, young female, elderly male, and elderly female) listened
to the vowel stimuli originally produced by 40 speakers (20 young and 20 old speakers), and categorized the age of each speaker into two age groups: young (25 years old or younger) or old (75 years old or older). They employed the three kinds of vowel stimuli, phonated vowels, whispered vowels, and low-pass filtered vowels, but the main interest in this study was the correct identification results for the phonated vowels. The elderly listeners were less accurate judging the age of elderly speakers than the younger listeners (correct age identifications of elderly speakers were 67.5% in the young listeners and 53.3% in the elderly listeners). Moreover, the elderly listeners showed that their age identification of the elderly speakers (53.3%) was less accurate than their identification of the young speakers (61%). Jacques and Rastatter suggested that the elderly listeners did not perceive the age-related acoustic properties that the young listeners were able to perceive to identify the age of elderly speakers. Obviously, noticeable inaccuracy among elderly listeners in Jacques & Rastatter (1990) could be partially explained by the use of more restricted speech samples (sustained vowels) than other studies. Because the main purpose of their study was to examine acoustic properties that result from age-related changes in the vocal tract, they discussed inaccurate age identification within the elderly groups from the acoustic point of view, but did not consider the other possibilities that Huntley et al. (1987) suggested.

Hearing sensitivity can certainly contribute to declined accuracy of age identification among the elderly listeners. However, hearing sensitivity is usually more impaired in male population than females. If the elderly listeners were not able to perceive some acoustic features due to age-related hearing impairment, it could be possible to find a difference in accuracy between male and female listeners in the elderly
group. However, both elderly male and female listeners demonstrated about the same accuracy in age identification (Jacques & Rastatter, 1990).

As suggested in Huntley et al. (1987), if lack of social contact was the main reason for older listeners’ poor performance, then the concepts of vocal age should change or deteriorate without constant exposure. This is an interesting possibility but there is no research addressing this issue so far.

Large age differences between the younger speakers and the older listeners were listed as another possible reason for elderly listeners’ overestimation of the age of younger speakers in Huntley et al. (1987). This explanation is similar to the ones provided to account for younger listeners’ underestimation toward older voices, except in the opposite direction. For older listeners, the young category might be widened such that discrimination of younger voices is coarse. This suggests that besides perceived age and chronological age, the effect of listeners’ “conceptual age” should be considered. Listeners may differently conceptualize age categories. That is, one listener thinks that a 35-year-old person belongs to the same young age group as a 25-year-old person, but another listener thinks that a 35-year-old person is considerably older than a 25-year-old. Then, listener’s perceived age could be affected by their own categorization of age. The underestimation of older talkers by young listeners seems a general tendency across studies (e.g. Shipp & Hollien, 1969; Hollien & Tolhurst, 1978; Mulac & Giles, 1996). This can be explained by the age difference between the speakers and the listeners, because listeners were selected from the young population. In fact, middle-aged listeners indicated the best performance of age estimations among the four listener groups in Huntley et al. (1987). This result implies that people are good at estimating age of
speakers in their age bracket (Hollien & Tolhurst, 1978). However, Huntley et al. (1987) presented contradictory evidence that the older listeners actually underestimated the age of talkers in their own age group. One possible explanation is drawn from the remarks of Huntley et al. (1987), “The older group (mean age 70.0 years) would perceive most of the talkers as being younger than themselves”. It is possible that listeners tend to use a biased scale of age depending on their own chronological age.

**Listener sex**

Many investigators did not specify the gender of listeners who participated in age estimation studies. Only a few studies explored the relationship between listener sex and perceived age. Results were consistent across studies, showing that listener sex did not make much difference in the accuracy of age perception for speakers (Braun, 1996; Jacques & Rastatter, 1990; Mulac & Giles, 1996; Hartmann, 1979), except that Hartman (1979) reported that the female listeners were consistently more accurate than the male listeners for the male speakers beyond 50 years. The reason for this difference is not clear. Braun (1996) reported a slight difference between male and female listeners in terms of confidence in their judgments. Whether listener’s age estimation was good (seven or less years difference between perceived age and chronological age) or bad (more than seven years difference in age estimation), female listeners were less confident about their judgments than male listeners. The male-female difference in the self-rated confidence in their age estimation was only observed in the listener groups with phonetic training but not among the naïve listeners. Besides, the difference was small (average confidence rating was 2.84 among 8 male listeners and 2.21 among 4 female listeners) and the number of female listeners was also small. Therefore, more evidence is necessary
to decide whether female listeners are less certain about their age judgments than male listeners.

**Listener expertise**

Listener expertise appears to have little effect on age perception. Most listeners in previous studies were untrained listeners. Although the majority of listeners were assumed to be students who majored in speech sciences, most of the investigators mentioned that their listeners had no training in phonetics. Braun (1996) specifically addressed the question of whether the trained listeners are more accurate to estimate speakers’ age than non-trained listeners. She employed twelve expert listeners who had extensive experience in the forensic analysis of anonymous voices and nineteen university students who had no particular training in auditory phonetics (non-expert listeners). The listeners were asked to estimate the speakers’ age. They listened to speech produced by twenty male smokers and twenty male non-smokers who were between 25 and 59 years. All the speakers read the German version of *the North Wind and the Sun*. Braun found that there was no significant difference in age estimation accuracy between the two groups, although agreement among the listeners was better in the expert listener group than the non-experts. Braun (1996) also found that there was no significant difference in self-confidence levels between the trained and untrained listeners. However, she also observed that female trained listeners were less confident about their judgments than male trained listeners as mentioned above. Similarly, Neiman and Applegate (1990) found that listeners were certain of their judgments regardless of whether their estimates were correct or not.
Social factors

Mulac and Giles (1996) reported the only age perception study that directly examined the effect of social factors. They examined the relationship between the psychological state of speakers and their perceived age. In addition to the conventional age measure of chronological age, the relationship between listener’s perceived age and the two non-chronological ages (subjective age and contextual age measures) was investigated. Subjective age was determined by speakers’ self-judgments on their activities, feeling, appearance, and interests. Contextual age was also self-perceived age, but based on the amounts of physical activities, health conditions, mobility, social interactions, and psychological and economical satisfaction in their life. They predicted that the perceived age is correlated better with these non-chronological ages than chronological age. Young listeners listened to short conversational speech by 36 speakers aged between 59-92 years old, and estimated the age of speakers. Results revealed that the listeners’ perceived age was correlated best with the chronological age, \( r = 0.51 \). Correlation between perceived age and speaker’s subjective age was somewhat less, \( r = 0.45 \), and correlation between perceived age and contextual age was not significant. It was surprising that they did not find a significant correlation between contextual age and perceived age because the age perception is affected by the general state of speaker’s physiological condition (Ramig, 1983a, 1983b; Ramig & Ringel, 1983). The restricted age range of the talkers might be the reason that they failed to find the relationship between perceived age and contextual age. To evaluate the relationship between perceived age and contextual age, people with a higher contextual age than their chronological age should be included in the data. Since it is possible that listeners’
internal categories of age interacts with the age estimation, those non-chronological ages may be more useful to explore the nature of age measures that the listener uses. Significance of these socio-cultural, socio-psychological factors remains to be clarified.

Although their evidence was not direct, results of Brawn (1996) seemed to imply that the listeners could draw upon their socio-cultural knowledge, when they estimated the age of talkers. In Brawn (1996), the age of German speakers was estimated. Half of the speakers were non-smokers, and other half of the speakers were smokers. Age estimation errors observed for the non-smoking speakers were larger than estimation errors for the smoking speakers. In other words, listeners tend to underestimate non-smoking speakers’ age. Braun (1996) sought an explanation of her rather counter-intuitive result from speakers’ physiological condition. Since smoking causes fibrotic deterioration in the larynx as well as a decrease of respiratory functions, smokers appeared to be in poor physiological condition. Based on the finding that the listeners judge biological age rather than chronological age (Ramig, 1986), Braun drew a conclusion that this large mismatch between chronological age and biological (physiological) age among the non-smokers could cause listeners’ underestimation of speakers’ age. However, when the majority of male speakers were not smokers, a mismatch between chronological age and biological age should be larger for the smokers. Statistics of smoking prevalence provides another explanation for the result that the age of smokers was more accurately perceived than the age of non-smokers. Estimated prevalence of smoking among German males has been about 30% for the last decade. It is slightly higher than in the United States. However, historically, the percentage of male
smokers in Germany was 59% in 1971 and 88% in 1950\(^1\). Apparently, the majority of German males used to be smokers. In other words, if speakers are male, especially German male, smokers are more representative male speakers than non-smokers. Therefore, the result may indicate that listeners are sensitive to their demographic knowledge when they estimate the age of speakers.

The age range of perceived ages also implied that the listeners were sensitive to demographic information when they estimated the age of speakers. Unlike previous studies, Baker (1981) did not provide specific instructions on speakers’ age range to the listeners. Both groups of listeners applied an age range for their answers beyond the lower limit of speaker age but below the upper limit of speaker age. Mean perceived ages by both listeners for both sexes were approximately normally distributed within 20 to 79 years. Furthermore, mean perceived age for female speakers was higher than that for male speakers. These results strongly suggest that listeners apply their demographic knowledge in age perception.

Summary

Research has shown that listeners estimate speaker’s age quite accurately from speech alone. Extensive research has been conducted to discover the acoustic properties associated with speaker age. On the other hand, a relatively small number of perception studies have been conducted. The majority of research on age perception has been based on English. Although similar results were found for languages other than English, the effect of speaker’s native language on the age perception has not yet been examined. In

\(^1\) The data in 1950 is for West Germany. Other German data is for both West and East Germany. (Forey, Hamling, Lee, & Wald, 2002)
spite of the lack of cross-language comparison, apparently there is consensus among researchers that elderly speech undergoes the same changes no matter what the speaker’s background is. Since it is not clear whether or not elderly speech characteristics are consistent across various factors, this generalization about elderly speakers is premature. Although some researchers provided brief comments on possible socio-cultural effects, there has been very little attempt to examine the socio-cultural effects on elderly speech.

Some studies have implied that the listeners are sensitive to demographic information when they estimate the age of speakers. Socio-psychological perception of the elderly seems to play an important role in age perception as well.

The current study explores several factors that influence perception of age. In particular, the current study focuses on the effect of language familiarity, amounts of linguistic information, speaker’s sex, speaker’s age, and conventional age stereotypes on the listeners’ perception of age.
CHAPTER 2

2. Speech Corpus

A speech corpus, the Vocal Age Speech Corpus (VASC), was developed in this project to enable research on age-related changes in speech. The main purpose of the VASC was to collect speech data from talkers, including both men and women, with different language backgrounds and various ages. The current speech corpus includes speech from sixty speakers: they are divided in terms of language: English and Japanese; age groups: young, middle-aged, and elderly; and sex: male and female. The corpus consists of three types of speech material: pronunciation of sustained vowels, pronunciation of words embedded in a carrier sentence, and readings of short passages.

The targeted age ranges were 25-30 years for the young group, 55-60 years for the middle-aged group, and 80-85 years for the elderly group. The age range for the young group was selected in order to avoid those who are in the late stage of developmental changes in speech. Hence, the selected age range for the young group was slightly older than the average age of undergraduate students. The selection of age range for the middle-aged group was determined mainly to avoid women who are experiencing menopause, which could introduce a lot of individual variability in the female data. The average age of menopause is around 52 years old in the United States as well as in Japan, so that the lower age limit for the middle-aged group was set to 55 in the current project. The elderly group was selected to represent people in the late stages of life.
Participants

Recruitment and recording of participants were carried out in two locations. English talkers participated in Bloomington, Indiana, while Japanese talkers in Kobe (and the surrounding areas), Japan. Talkers in both language groups met the following criteria (in addition to the conditions on age and linguistic backgrounds described above).

All talkers had no past and present history of speech or language disorder. They also had no history of stroke or any other neurological disorders. They also had no history of surgery or injury of the laryngeal area, except three Japanese middle-aged talkers (JFM01, JMM04, and JMM05) who had adenoidectomies during their childhood.

Also, all had normal hearing (the average pure tone threshold for the frequencies of 500, 1000, 2000, and 4000 Hz < 35dB HL), although young and middle-aged talkers exhibited a better hearing sensitivity than many elderly talkers.

In addition, all talkers were ambulatory. Some of the elderly talkers preferred to use a walking stick, but they were also capable of walking without a stick. In addition to their walking ability, none had any problem or required assistance for daily activities such as bathing, dressing, toilet use, and eating.

English talkers

The Vocal Age Speech Corpus (VASC) includes 30 native speakers of American English. They were recorded in Bloomington, Indiana, from January through March, 2005. Some of them participated as volunteer subjects, and some were paid for their participation. In addition, there were eight other people who participated in the recording, but their speech was not included in the corpus due to various reasons such as poor audio
quality, unsatisfactory backgrounds, or incomplete data. The means and ranges of age for each group of English talkers at the time of their recording are summarized in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young</td>
<td>25.4(24-28)</td>
<td>26.2(24-30)</td>
<td>25.8(24-30)</td>
</tr>
<tr>
<td>Middle</td>
<td>57.8(56-60)</td>
<td>56.2(54-59)</td>
<td>57.0(54-60)</td>
</tr>
<tr>
<td>Elderly</td>
<td>82.2(80-83)</td>
<td>82.0(80-86)</td>
<td>82.1(80-86)</td>
</tr>
</tbody>
</table>

All of the English talkers were non-Hispanic white American\(^2\). Their present or past occupations were varied (see Appendix A1). No one had engaged in physically strenuous labor.

The majority of English speakers were native speakers of some variety of American English. Some of them, especially elderly speakers, were not originally from the Midwest. However, they had long residence history within Indiana or neighbor states, and did not exhibit a strong dialectal accent specific to a region different from Indiana.

Although most of the English talkers were monolingual speakers of American English, some were able to speak foreign language(s) such as French, Spanish, and Chinese. However, these foreign languages were not used as the main language in their daily life. More importantly, none of English talkers were able to speak any Japanese.

None had received professional training in public speaking. A few talkers had received formal training in singing, and one talker had a formal training in acting. Detailed information about individual talkers can be found in Appendix A1.

\(^2\) Ethnicities were not included in the criteria for selecting English talkers.
The number of smokers was surprisingly small in English talkers. However, about half of middle-aged and elderly men and middle-aged women used to be tobacco smokers.

Table 6 summarizes experiences with smoking.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age</th>
<th>Ex-smoker</th>
<th>Smoker</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Young</td>
<td>6.0</td>
<td>na</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>(N=1)</td>
<td>(N=0)</td>
<td></td>
<td>(N=1)</td>
</tr>
<tr>
<td></td>
<td>Middle-aged</td>
<td>7.0</td>
<td>na</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>(N=2)</td>
<td>(N=0)</td>
<td></td>
<td>(N=2)</td>
</tr>
<tr>
<td></td>
<td>Elderly</td>
<td>15.0</td>
<td>na</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>(N=3)</td>
<td>(N=0)</td>
<td></td>
<td>(N=3)</td>
</tr>
<tr>
<td>Female total</td>
<td></td>
<td>10.8</td>
<td>na</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>(N=6)</td>
<td>(N=0)</td>
<td></td>
<td>(N=6)</td>
</tr>
<tr>
<td>Male</td>
<td>Young</td>
<td>9.0</td>
<td>na</td>
<td>9.0</td>
</tr>
<tr>
<td></td>
<td>(N=1)</td>
<td>(N=0)</td>
<td></td>
<td>(N=1)</td>
</tr>
<tr>
<td></td>
<td>Middle-aged</td>
<td>25.0</td>
<td>29.0</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td>(N=2)</td>
<td>(N=1)</td>
<td></td>
<td>(N=3)</td>
</tr>
<tr>
<td></td>
<td>Elderly</td>
<td>16.3</td>
<td>na</td>
<td>16.3</td>
</tr>
<tr>
<td></td>
<td>(N=3)</td>
<td>(N=0)</td>
<td></td>
<td>(N=3)</td>
</tr>
<tr>
<td>Male total</td>
<td></td>
<td>18.0</td>
<td>29.0</td>
<td>19.6</td>
</tr>
<tr>
<td></td>
<td>(N=6)</td>
<td>(N=1)</td>
<td></td>
<td>(N=7)</td>
</tr>
</tbody>
</table>

**Japanese talkers**

Thirty native speakers of Japanese were recorded from November through December, 2004, in Kobe and the surrounding areas, Japan. Most were paid for their participation, but some participated as volunteers. In addition, thirteen other Japanese speakers participated in the recording, but their data were not included in the corpus due to poor sound quality, incomplete data, or they did not meet some of the requirements set for Japanese speakers. The means and ranges of age for each group of Japanese talkers at the time of their recording are summarized in Table 7.
Table 7. Mean and range of age of Japanese speakers.

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young</td>
<td>27.0(25-29)</td>
<td>28.2(26-30)</td>
<td>27.6(25-30)</td>
</tr>
<tr>
<td>Middle</td>
<td>56.8(55-58)</td>
<td>56.6(55-60)</td>
<td>56.7(54-60)</td>
</tr>
<tr>
<td>Elderly</td>
<td>82.2(80-86)</td>
<td>82.8(80-86)</td>
<td>82.5(80-86)</td>
</tr>
</tbody>
</table>

All Japanese talkers had lived in Kobe and the surrounding areas for a long time. The native dialects were not the same across all talkers. However, their dialects were nevertheless a variety of the Kansai dialect or a dialect that has the same word accent type as the Kansai accent type. Hence, talkers did not exhibit a noticeable accent that deviated from the Kansai dialect.

All were virtually monolingual speakers of Japanese, although English has been introduced as a mandatory subject in Japan. Each talker’s oral proficiency of English was extremely limited, because attention has been almost exclusively focused on achievement of literacy-related abilities in Japanese school system. In addition, the speakers had not spoken English since they finished schooling.

No one had received formal training in public speaking and acting. A few of elderly talkers were practicing shigin singing (recitation of Chinese poems) for entertainment for their old age. One female elderly talker reported that she had professional singing training as a member of a church choir. Detailed information about individual talkers can be found in Appendix A2.

Compared with the English talkers, the number of female smokers (including former smokers) was about same except that there is a clear difference in the elderly females. Japanese elderly females had never had a smoking habit in their life, but three
out of five English elderly females had a smoking habit when they were young. On the other hand, there were more current and former tobacco smokers among Japanese men. In particular, the Japanese elderly men had a longer duration of smoking than the English elderly men. Table 8 summarizes the average years of smoking for Japanese talkers.

Table 8. Average years of smoking for Japanese present and former smokers.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age</th>
<th>Ex-smoker</th>
<th>Smoker</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>na</td>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>Female</td>
<td>Young</td>
<td>(N=0)</td>
<td>(N=1)</td>
<td>(N=1)</td>
</tr>
<tr>
<td></td>
<td>Middle-aged</td>
<td>(N=2)</td>
<td>na</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>Elderly</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>Female total</td>
<td>15.0</td>
<td>7.0</td>
<td>12.3</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Young</td>
<td>(N=2)</td>
<td>(N=1)</td>
<td>(N=3)</td>
</tr>
<tr>
<td></td>
<td>Middle-aged</td>
<td>(N=2)</td>
<td>35.0</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td>Elderly</td>
<td>(N=4)</td>
<td>na</td>
<td>42.5</td>
</tr>
<tr>
<td>Male total</td>
<td>26.0</td>
<td>22.0</td>
<td>25.2</td>
<td></td>
</tr>
</tbody>
</table>

Speech materials

Three types of speech were collected from each speaker in his/ her own native language. These types were sustained phonation of vowels, read sentences, and read passages.

Sustained vowels were collected in order to evaluate voice characteristics in general. Also, the data can be used for comparison with previous and possibly future research as well because sustained vowels have been widely used for studies of voice
characteristics. The sustained vowels included two vowels (/a/ and /i/). Each vowel was produced for about five seconds as steadily and clearly as possible in speaker’s comfortable speaking voice. Also, each vowel was pronounced at three pitch levels: comfortable pitch level, high pitch level, and low pitch level. There are three repetitions for each vowel at each pitch level.

There were two reasons for recording passages. One was to record standardized texts that have been used in various fields of studies on speech. The other purpose was to obtain speech that contained a variety of linguistic and speech characteristics. Thus the VASC includes two standardized texts. “The Rainbow Passage” (Fairbanks, 1960) is selected because of its extensive use in the United States. It was also selected in order to compare the outcome from the current data with the results of those studies. However, because “The Rainbow Passage” is not commonly used in Japan, it was translated by two native speakers of Japanese to use for monolingual Japanese talkers. The VASC also includes reading of “The North Wind and the Sun” or “Kitakaze to Taiyo” (the Japanese version of “The North Wind and the Sun”), which has been used in many cross-language studies because the texts are available with IPA transcriptions in various languages (International Phonetic Association, 1949). The reading of “The North Wind and the Sun” was selected in order to expand the current paradigm to other languages for future research.

Finally, the VASC includes sentences read in various conditions. This part of the corpus was specially designed for the present study. Essentially sustained vowels have too little information while the passage contains too much information. Although each passage includes the same content in both languages, cross-language comparisons would
be extremely difficult because of the differences in linguistic structures. In order to
examine speech characteristics across languages, more controlled speech material is
desirable. The sentence part of the corpus was designed to fulfill these conditions.

One target sentence was read in various conditions. The sentence includes the
target phrase, which is composed of a sequence of alphabet letters, “BCC”. This phrase
was chosen because it is a possible word in both languages, and acoustically similar in
both languages, /bisisi/. The target phrase was embedded in a carrier sentence. In other
words, the target sentence served as a carrier sentence for the target phrase. The English
carrier sentence is shown in (1), and the Japanese one in (2).

(1) The answer was ___ Corporation.
(2) Kotae wa _____ kooporeeshyon datta yo. ‘The answer was ___ Corporation.’

The sentence part of the corpus included sentences produced in four conditions:
normal condition, age-disguise condition, impersonating condition, and dialogue
condition. The main aim of the age-disguise and impersonating conditions was to obtain
stereotyped speech associated with age or a certain age group.

In the normal condition, sentences were read in a normal fashion. In the age-
disguise condition, speakers were asked to read the sentence as if they were pretending to
be either 20 years older or 20 years younger. Hereafter, these two sub-conditions are
termed the maturation condition and the rejuvenation condition, respectively.
In the impersonating condition, speakers were asked to read the same sentence pretending to be a specific speaker of the sentence. They were asked to pretend to be a baby, a mother, a father, a grandmother, and a grandfather.

Finally, the VASC included sentences in the dialogue condition. They were collected to obtain utterances with or without corrective focus. Using the target sentence described in (1) and (2) above, the sentences were read as an answer responding to a question as if it were a dialogue. The dialogues were designed to elicit either broad focus or corrective focus for the target phrase by changing the form of the question.

Summarizing, the corpus includes three repetitions of each of two sustained vowels at each of three pitch levels. There were three repetitions of the target sentence in every sentence condition, and two readings of each passage.

**Recording procedures**

Audio data were recorded to a hard disk of a laptop computer Sharp Mebius PC-GP1-C5U at 44.1 kHz sampling frequency with 16 bits per sample, using Cool Edit 2000. A head-set microphone Shure SM10A was used with a USB-powered preamplifier M-Audio MobilePre USB. Recording was conducted in quiet to moderately quiet environments such as laboratory rooms or small classrooms in schools, reception rooms in offices or community centers, or living rooms in the participant’s own home.

Participants were provided a glass of water with a drinking straw and advised to drink water any time they wished during the recording.

Each speaker was first asked to read and sign a consent form. Then, they were asked to fill out a questionnaire. The questionnaire was composed of four sections: 1)
background information, 2) physical activities, 3) dietary and social habits, and 4) past and present health status. The questionnaire can be found in Appendix B. Background information included demographic information such as race, sex, age, education, occupation, linguistic information, and residential history and environments.

The section about physical activities asked about frequency and intensity of physical activities that the person engaged in the past and present daily life. It also included the question of independence of the person in terms of daily activities. Physical activity levels of daily living were used for an indirect measure of physiological conditions of each speaker. It has been reported that speaker’s general status of physiological conditions affects voice characteristics (Ramig, 1983a, 1983b, 1986; Ramig & Ringel, 1983; Ringel & Chodzko-Zajko, 1987). In general, if speakers are in a poor physiological condition, their voices tend to sound more breathy, shaky, or coarse than the voice produced by speakers in a good physiological condition. Speech rate and intonation can be different. The questions in this section were created based on the Index of Activities of Daily Living Scale (Ainsworth, Jacobs Jr., & Leon, 1993; Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963), and the guideline provided by the National Center for Disease Control and Prevention (August 23, 2003).

The section asking about the speaker’s dietary and social habits was designed to obtain information about past and present consumption of tobacco, alcohol, and caffeine. It is known that tobacco, alcohol, and beverages containing caffeine affect the characteristics of speech. Especially, smokers tend to have more breathy and hoarse voice quality than non-smokers (Braun & Rietveld, 1995). Excessive consumption of alcoholic beverages and of beverages containing caffeine have a dehydrating effect,
which often makes the mouth dry. This can cause the voice to sound more breathy and hoarse.

The last section on health status was designed to collect various health conditions that could affect speech. These included a history of speech, language, or hearing problems, a history of strokes or some neurological diseases, dental information, a history of medicines that may create either a drying effect or a reduction in glandular secretions.

After filling out the questionnaire, the experimenter placed the Shure SM10A head-set microphone on the speaker so that there was about 1 inch distance between the microphone and the speaker’s mouth. Recording of each talker proceeded in the same order as shown in Table 9 and took about one hour. Instructions for each condition are found in Appendix C.

<table>
<thead>
<tr>
<th>Recording order</th>
<th>Speech type</th>
<th>Condition</th>
<th>Number of Repetitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vowel</td>
<td>Normal pitch level</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Vowel</td>
<td>High pitch level</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Vowel</td>
<td>Low pitch level</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Sentence</td>
<td>Normal condition</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>Sentence</td>
<td>Dialogue condition</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>Sentence</td>
<td>Age-disguise condition</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>Sentence</td>
<td>Impersonated condition</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>Passage</td>
<td>“The North Wind and the Sun”</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>Passage</td>
<td>“The Rainbow Passage”</td>
<td>2</td>
</tr>
</tbody>
</table>

First, each speaker was asked to pronounce vowels (/a/ and /i/) in a comfortable speaking voice and sustain their pronunciation. They were instructed to pronounce the vowel as steadily and clearly as possible, keeping the same pitch and loudness levels throughout each pronunciation. They were told to sustain each pronunciation
approximately five seconds until the experimenter indicated they should stop. Each talker started with /a/, and continued with /i/. For English talkers, vowels were specified with a conventional phonetic transcription with example word (e.g. “ah” as in “ba-ba black sheep” and “ee” as in “bee keeper”). They were asked to repeat each vowel three times in the same fashion. Then, they were asked to pronounce the same vowels in the same way but at a high pitch level that was still comfortable. After that, they were asked to do the same thing at a low pitch level.

After these vowel conditions, speakers continued to record the sentence conditions in the order specified in Table 9. In each condition, speakers were asked to read sentences written on cards. All text on the cards was written in an 18-point font size for better visibility.

In the normal sentence condition, speakers were instructed to simply read the sentence written on a card. Seven similar sentences were used as distractors for the target sentence. All eight sentences are found in Appendix C. The speakers were given a list of the eight sentences to read prior to the recording. They were told that they were going to read each sentence in random order three times as shown on a card provided by the experimenter. The speakers were instructed to read all sentences in the same way. When the speaker stumbled, they were asked to read the same sentence again.

In the dialogue condition, speakers were asked to read a short dialogue between two people. They were instructed to imagine a conversation between two people in the following situation:

“Two people (Pat and Terry) are watching a popular quiz show on TV. Today’s show has various questions about the famous companies. In
the last question, the respondents were asked, “Which one of the companies remains atop the industry?” One contestant gave an answer. But, Pat didn’t catch the answer for this question. Pat will ask Terry about the answer. Terry will tell Pat the answer.” [The Japanese version uses Yamada-san and Suzuki-san instead of Pat and Terry.]

The English speakers were given examples of a short dialogue as seen in (3) and (4). The dialogue in (3) was designed to elicit broad focus, while the one in (4) was to elicit corrective focus on the second letter in the target phrase. The corresponding Japanese dialogues seen in (5) and (6) were given to the Japanese speakers.

(3) Pat: What was the answer?
   Terry: The answer was BCC Corporation.

(4) Pat: Was it BBC?
   Terry: No, the answer was BCC Corporation.

(5) Yamada: Kotae-wa nan datta?
   Suzuki: Kotae-wa biisiisii kooporeesyon datta yo.

(6) Yamada: Kotae-wa biibiisii kooporeesyon datta?
   Suzuki: Iya, kotae-wa biisiisii kooporeesyon datta yo.

In addition to one interrogative sentence designed to elicit correct focus on the second letter in the target phrase BCC as in (4) or (6), two interrogative sentences similar to these were prepared to elicit corrective focus on the first and the third letter in the target phrase. Each of these four dialogues was written on a separate card. Each speaker was asked to read each dialogue three times in a random order.

In the next age-disguise condition, speakers were asked to read the target sentence as if they were 20-years younger or 20-years older than their age. Each card had the
target sentence written with an instruction of pretended age such as “20 years younger” or “20 years older”. The talkers had time for practice before their recordings. The sentence was read three times for each disguised age in a random order.

In the impersonating condition, speakers were given a list of five sentences as in (7) through (11) below. As you can see, these sentences include the target sentence as a quote, but the speaker is either a baby bear, a mother bear, a father bear, a grandmother bear, or a grandfather bear. The participants were asked to read a whole sentence written on a card, but read only the part within the quotations as if they were the speaker of the sentence. Practice time was given to each participant. The five sentences were blocked and presented three times. The order of the five impersonated speakers was randomized in each block.

(7) When Baby Bear talks, he sounds like this,

“The answer was BCC Corporation.”

(8) When Mama Bear talks, she sounds like this,

“The answer was BCC Corporation.”

(9) When Papa Bear talks, he sounds like this,

“The answer was BCC Corporation.”

(10) When Grandma Bear talks, she sounds like this,

“The answer was BCC Corporation.”

(11) When Grandpa Bear talks, he sounds like this,

“The answer was BCC Corporation.”

Finally, the speakers were given a sheet that had “The North Wind and the Sun” written in their native language. After the speakers carefully read the passage in silence,
they were asked to read the passage aloud two times. Then, a sheet of “The Rainbow Passage” written in their native language was given to the speakers. Then again, they were asked to read the passage carefully first in silence, and then to read it aloud twice. They were instructed to continue to read each passage until the end even if they missed a word. See Appendix C for the exact instructions. An 18-point font size was used for the texts.

Immediately after the recording, speakers were asked to take a hearing test. The hearing test was conducted using AUDIO-CD (Digital Recordings, 1997). The frequencies used for this project were 500, 1000, 2000, and 4000 Hz. Each speaker tested the left ear before the right ear. They were asked to raise their hand when they heard a sound. All the tones were played by the Windows Media Player 9, through a stereo headphone Sony MDR-Z300.
CHAPTER 3

3. The effects of language familiarity and the amount of information

Introduction

A number of studies have shown that listeners’ estimation of a speaker’s age is fairly accurate (Ptacek & Sander, 1966; Shipp & Hollien, 1969; Ryan & Burk, 1974; Horii & Ryan, 1981; Neiman & Applegate, 1990; Braun, 1996; Cerrato et al., 2000; Baker, 1981a). Although various factors in age estimation have been examined, effects of language on age perception have not yet been explored. It seems that previous research on aging speech assumes that age-related speech characteristics are language independent. If this is the case, listeners should be able to estimate a talker’s age regardless of the talker’s language. However, this prediction has not been tested.

Correlations between actual and estimated age reported in this research are summarized in Table 10. However, except for a few studies on non-English languages (Braun (1996) for German; Cerrato et al. (2000) for Italian; and Kido & Kasuya (2004) for Japanese), the literature on age perception has been dominated by a single language, American English. Because of some methodological differences among the different researchers, we should be careful in interpreting these results. Still, comparing these studies suggests that listeners’ ability to estimate the age of speakers appears to be rather similar across languages.
Table 10. Correlation values between perceived age and chronological age in various studies.

<table>
<thead>
<tr>
<th>Source</th>
<th>Correlations</th>
<th>Age range of speakers</th>
<th>Speaker language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shipp &amp; Hollien (1969)</td>
<td>0.88</td>
<td>20-89</td>
<td>AE</td>
</tr>
<tr>
<td>Ryan &amp; Burk (1974)</td>
<td>0.77*</td>
<td>40-80</td>
<td>AE</td>
</tr>
<tr>
<td>Horii &amp; Ryan (1981)</td>
<td>0.76</td>
<td>40-80</td>
<td>AE</td>
</tr>
<tr>
<td>Baker (1981)</td>
<td>0.68* (Caucasian listeners)</td>
<td>40-95</td>
<td>AE</td>
</tr>
<tr>
<td></td>
<td>0.69 (African-American listeners)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramig et al. (1985)</td>
<td>0.17</td>
<td>25-75</td>
<td>AE</td>
</tr>
<tr>
<td>Neiman &amp; Applegate (1990)</td>
<td>0.91</td>
<td>20-75</td>
<td>AE</td>
</tr>
<tr>
<td>Braun (1996)</td>
<td>0.68 (untrained listeners)</td>
<td>25-59</td>
<td>German</td>
</tr>
<tr>
<td></td>
<td>0.70 (trained listeners)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cerrato et al. (2000)</td>
<td>0.77</td>
<td>18-66</td>
<td>Italian</td>
</tr>
<tr>
<td>Kido &amp; Kasuya (2004)</td>
<td>0.66</td>
<td>20-60</td>
<td>Japanese</td>
</tr>
</tbody>
</table>

* Correlation values were calculated based on the data by the author. AE stands for American English.

Besides various factors in the acoustic signal, the literature suggests that factors that are not available in speech signals also affect age perception. These factors include listener age (Huntley et al., 1987), listener sex (Braun, 1996; Hartmann, 1979; Kukol, 1979), speaker’s smoking habits (Braun & Rietveld, 1995), and ethnicity (Baker, 1981). In particular, the effect of listener age has been discussed repeatedly. It is common that listeners’ age estimation is more accurate for talkers whose ages are close to the listener’s. Shipp and Hollien (1969) gave an interpretation of this as a peer effect, suggesting that familiarity with individual voices could affect age perception. Similarly, Baker (1981) found that Caucasian listeners estimated the age of African-American male speakers as younger than African-American listeners did. She suggested that this difference was related to the listeners’ different experiences with older people in the past.
The current study explored the effect of language familiarity on the perception of talkers’ age. The two languages compared are English and Japanese. English was selected because the results will enable a direct comparison with current literature. Japanese was chosen for a cross-comparison for several reasons. One reason is that Japanese is not a familiar language for English listeners. Another reason is related to the fact that Japan has the longest average life expectancy in the world. Average life expectancy for Japanese men and women in 2004 were 78.64 and 85.59 years old, respectively (Japan Ministry of Health, Labour, and Welfare, 2004). Average life expectancy in the United States in 2002 was similarly long, 74.5 years for men and 79.9 years for women (U.S. Department of Health and Human Services, 2004). Hence, although Japanese has longer life expectancy than American, listeners in both countries could have a similar demographic knowledge in terms of the age distributions.

There are several interesting differences between the Japanese and American experience with aging, which may result in asymmetrical estimations of age between the two language groups. Although average life expectancies (from birth) in both countries are similarly long, the living arrangements of elderly people are quite different in these countries. A high percentage of Japanese elderly co-reside with their children. This is not a common living arrangement for elderly in the developed countries including the United States. Brown et al. (2002) reported that living with children has been the most stable living arrangement in Japan during the period of their survey from 1987 to 1996. They surveyed living arrangements of 4878 Japanese elderly whose ages were over 60 and had at least one living child in 1996, and found that 52% of Japanese elderly lived with their children. Although the number had dropped about 18% from 1980 to 1996,
since 1996, the percentage of elderly (over 60 years) who co-reside with their children remained about same, 53.8% (Cabinet Office, 2001). This is certainly a high percentage compared with only about 15% of elderly living with their children in the United States, which is one third of the Japanese rate (Cabinet Office, 2001). In addition, about one fourth of Japanese elderly live with their grandchildren, while only 5.3% of American elderly do (Cabinet Office, 2001). These differences in living arrangements do not necessarily mean that the American younger adults do not interact with the elderly people; however, in general young Japanese adults have more opportunities to have face to face communicative interactions with the elderly, and this increased exposure may result in better age estimation abilities in young Japanese listeners.

In addition to the effect of language familiarity, the current study examines the effect of the degree of linguistic information on listener’s age estimation. A majority of age perception research employed a stimulus containing ample linguistic information. This type of the stimulus included some kind of read speech (Ptacek & Sander, 1966; Shipp & Hollien, 1969; Ryan & Burk, 1974; Neiman & Applegate, 1990; Braun, 1996; Cerrato et al., 2000; Baker, 1981a; Kukol, 1979; Ramig, 1986; Horii & Ryan, 1981), or extemporaneous speech (Hartmann, 1979; Mulac & Giles, 1996; Benjamin, G. R., 1992). On the other hand, a few studies have used sustained vowels as their stimuli (Ptacek & Sander, 1966; Ramig et al., 1985; Linville & Korabic, 1986; Linville & Fisher, 1985b), which include little linguistic information. Types of the stimuli used in previous studies and the obtained correlations values between perceived age and chronological age are listed in Table 11.
If listeners can judge the age of a speaker based on voice characteristics, vowel-only stimuli would provide sufficient information to estimate the speaker’s age. However, it seems that the listeners need linguistic information in addition to voice information to estimate the speaker’s age. When listeners estimated the age of talkers from sustained pronunciations of the vowel /a/, listeners correctly identified the age group of talkers at higher levels of performance than a chance level when the identification was for two age groups (i.e., young vs. old) (Ptacek & Sander, 1966) or for three groups (i.e., young, middle-age, and old) (Linville & Fisher, 1985b; Linville & Korabic, 1986). However, when the listeners used five age categories for judging the vowel stimuli, the correlation between speakers’ chronological age and perceived age was extremely low, $r = 0.17$ (Ramig et al., 1985). It is also known that speaker identification is more accurate when longer speech samples are presented to the listeners (Pollack, Pickett, & Sumby, 1954; Compton, 1963).
Table 11. Correlation values between perceived age and chronological age in various studies.

<table>
<thead>
<tr>
<th>Source</th>
<th>Stimuli</th>
<th>Correlations</th>
<th>Age range of speakers</th>
<th>Speaker language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ramig et al. (1985)</td>
<td>1 vowel</td>
<td>0.17</td>
<td>25-75</td>
<td>AE</td>
</tr>
<tr>
<td>Kido &amp; Kasuya (2004)</td>
<td>1 sentence</td>
<td>0.66</td>
<td>20-60</td>
<td>Japanese</td>
</tr>
<tr>
<td>Shipp &amp; Hollien (1969)</td>
<td>1 sentence</td>
<td>0.88</td>
<td>20-89</td>
<td>AE</td>
</tr>
<tr>
<td>Ryan &amp; Burk (1974)</td>
<td>2 sentences</td>
<td>0.77*</td>
<td>40-80</td>
<td>AE</td>
</tr>
<tr>
<td>Baker (1981)</td>
<td>1 paragraph</td>
<td>0.68* (Caucasian listeners) 0.69 (African-American listeners)</td>
<td>40-95</td>
<td>AE</td>
</tr>
<tr>
<td>Neiman &amp; Applegate (1990)</td>
<td>3 sentences</td>
<td>0.91</td>
<td>20-75</td>
<td>AE</td>
</tr>
<tr>
<td>Braun (1996)</td>
<td>1 paragraph</td>
<td>0.68 (untrained listeners) 0.70 (trained listeners)</td>
<td>25-59</td>
<td>German</td>
</tr>
<tr>
<td>Cerrato et al. (2000)</td>
<td>40 second passage</td>
<td>0.77</td>
<td>18-66</td>
<td>Italian</td>
</tr>
</tbody>
</table>

* Correlation values were calculated based on the data by the author.

Research on age perception conducted in the United States frequently employed a standardized text of *The Rainbow Passage* (Fairbanks, 1960), although there are slight differences among the studies as to what part of *The Rainbow Passage* was selected. For example, Shipp & Hollien (1969) used the third sentence of the passage, while Ryan & Burk (1974) used the fourth and the fifth sentences. Neiman and Applegate (1990) used first three sentences for their stimuli. Kukol (1979) and Baker (1981) used the entire opening paragraph of the passage, while Ramig (1986) used the first part of the third sentence. In general, the longer a stimulus is, the more information it contains as well as more variability. Hence, it is very difficult to specify exactly what kinds of information were used by the listeners.
Among the age estimation studies, Neiman and Applegate (1990) is the only study that used two types of stimuli in which the degree of linguistic information differed. In addition to the read speech of the rainbow passage, Neiman and Applegate (1990) employed phrase level material. They reported that the listeners’ judgments tend to be more reliable when they listened to the phrase stimuli than the sentence stimuli. Unfortunately, because the aim of their research was not the comparison between the speech materials, they only reported the reliability of judgments in terms of the types of speech materials. Even so, similar but slightly different results were found for speech material shorter than a paragraph suggesting that the difference in the amount of information could have a significant influence on the perception of age.

**Purpose**

The main purpose of this study is to explore the role of listeners’ language familiarity on their perception of a talker’s age. Two groups of listeners whose native languages are different (i.e. English and Japanese listeners) listened to speech from those who spoke in either a familiar or a less familiar language to the listeners (i.e. English and Japanese talkers). If language familiarity has any influence on listeners’ age judgments, results will show a significant difference in the perceived age whether listeners are listening to their native language or not.

This study also aimed to examine whether the different amount and kind of information could change listeners’ age estimations, and furthermore to examine whether language familiarity affects age perception for different linguistic contexts or not. For these purposes, three types of stimuli were used: 1) Vowel stimuli, sustained
pronunciations of /i/, which included the vocal characteristics of the talkers only; 2) Phrase stimuli, pronunciation of the letter sequence “BCC” extracted from dialogues, which included articulation dynamics and prosodic information in addition to the voice characteristics, and 3) Sentence stimuli, productions of one long sentence, which included more rich and detailed material and additionally contained much more language-specific information.

In order to see the language familiarity effect on age perception, data were grouped by whether the native language of the listener matched the one of the talker or not. Familiar language perception results are the perception results for talkers whose native language matches the listener’s, while less-familiar language results are the perception results for talkers whose native language is foreign to the listener.

There are several possible outcomes, five of which are considered here. Possible outcomes of the relationship between chronological age and perceived age are expressed relationally and described below (F: Familiar language data, L: Less-familiar language data):

(1) Vowel (L=F) = Phrase (L=F) = Sentence (L=F)
(2) Vowel (L=F) < Phrase (L=F) < Sentence (L=F)
(3) Vowel (L=F) ≤ Sentence (L) ≤ Phrase (L=F) < Sentence (F), where Vowel < Phrase
(4) Vowel (L<F) < Phrase (L<F) < Sentence (L<F)
(5) Vowel (L=F) < Phrase (L<F) < Sentence (L<F)
(1) If age perception is based solely on the voice characteristics, there will be no difference in the perceived age among the three contexts, or between the familiar and less-familiar data. If the amount of raw information influences age perception, differences among the three types of stimuli will be observed. (2) If more information helps the listeners to estimate the talker’s age, listeners’ age estimation will be improved from the vowel to the phrase to the sentence stimuli. Specifically, if the age of talkers indicated by certain acoustic properties is language-independent, perceived age will be the same whether the talkers speak in a familiar language or less-familiar language. (3) Alternatively, even if listeners use the same acoustic information to estimate age, judging foreign speech without understanding it may impair performance; hence, there will be a decline in listeners’ age estimation only for less-familiar language data at the sentence level. (4) Instead, if perception of talker age is language dependent, judgments for the talkers of a more familiar language will be better than judgments for those who speak a less-familiar language. (5) A more complicated version of this is that if the vowels are acoustically similar in both languages, then there will be no difference in terms of language familiarity at the vowel level (5). We expect that the relationships between chronological age (CA) and perceived age (PA) becomes stronger when listeners hear speech that has more information in it. Also stronger relationships are expected to be found in the familiar language data than less-familiar language data. However, the familiarity effect may not appear at the vowel level due to the acoustic similarity of the /i/ vowel in both languages.
Methods

Participants

Twenty-four native speakers of American English and twenty-four native speakers of Japanese participated in the perception experiment. In addition to those forty-eight listeners, eight other native speakers of American English participated in the experiment. However, their data were excluded from further analysis due to various reasons. The data from one participant was incomplete due to a technical problem. Another participant’s data was excluded because of the history of co-residency with a Japanese person in the past. The data from six other participants were excluded because four of them had extensive knowledge or use of other language(s), and two others grew up in different regions from the other participants.

English listeners were recruited and participated in the perception experiment in Bloomington, Indiana. They were students in various departments of Indiana University. All of them were native speakers of American English, had a long residence history in the Midwest (Nineteen participants were from Indiana, four from Ohio, and one from Illinois), and had no knowledge of Japanese language.

Japanese listeners were recruited and participated in the experiment in Kobe, Japan. They were students in Kobe Women’s College. All of them were native speakers of a Kansai dialect. Since English is a mandatory subject in Japanese schools, all of them had studied English for six years in their high school. However, none reported that they had any oral communication proficiency in English, which was indeed quite typical among Japanese in 2005. In a debriefing after the experiment, even though every Japanese listener reported that they identified the language they had heard as English,
none of them were able to understand the meaning of the sentences that they had repeatedly heard during the experiment. Both English and Japanese listeners reported that they had no hearing, speech, language, or neurological problems.

Means and ranges of age for each listener group were summarized in Table 12 and Table 13 below.

<table>
<thead>
<tr>
<th>Table 12. Mean age and age range of English listeners.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Range</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 13. Mean age and age range of Japanese listeners.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Range</td>
</tr>
</tbody>
</table>

**Stimuli**

Three kinds of stimuli were prepared from the speech corpus described above. Those three types will be called *vowel stimuli, phrase stimuli,* and *sentence stimuli.* Vowel stimuli contained 500 milliseconds of /i/ pronunciation. The vowel /i/ was chosen because of its acoustic similarities in English and Japanese. Phrase stimuli included just the letter sequence, *BCC,* edited from a carrier sentence (see Chapter 2 for detailed information). This phrase was chosen because it is pronounced similarly in both languages. Thus, phrase stimuli included speech dynamics without being radically different for the two languages. Sentence stimuli were taken from a portion of the reading of “The North Wind and the Sun” or “Kitakaze to Taiyoo,” which is often used in
speech science studies. The second sentence was selected (“They agreed that the one who could make the traveler take his coat off would be considered stronger than the other one” in English and “tabibito no gaitoo o nugaseta hoo ga kachi to iukotoni kimete mazu kitakaze kara hajime mashita.” in Japanese). The sentence obviously provided the listeners with a longer and more dynamic utterance. Because there were differences among talkers, such as duration of pauses, frequency of pauses, and speech rate, the sentences might indicate a speed of speech processing, which generally slows down with age.

Each vowel stimulus contained one second of sustained /i/ vowel from each talker. The portion selected was from a stable portion of each talker’s second /i/ pronunciation excluding the initial 500 milliseconds of pronunciation. The total number of the vowel stimuli was 60 (2 languages × 10 speakers × 3 age groups).

The phrase stimuli were prepared from sentences produced in the normal condition. Only the portion of BCC in the second repetition of the sentence was used for the stimuli. The phrase was edited from the end of the preceding phrase to the end of the last (third) vowel of the target portion BCC, as determined from waveforms and spectrograms. Thus, 60 phrase stimuli were prepared from sentences in the normal condition. In addition, 80 phrase stimuli were taken from the utterances in age-disguised (maturation and rejuvenation) conditions and 180 phrase stimuli were prepared from the sentences in the impersonated condition. The results for these phrase stimuli are not analyzed here, but will be presented in chapter 5.

For the sentence stimuli, the second sentence of “The North Wind and the Sun” was extracted from each talker’s first reading. If the talker made a speech error or a
correction, then the second reading was used instead of the first one. Total number of the sentence stimuli was 60.

Cool Edit 2000 was used to select and edit all speech materials. Each sound selection was cut at a zero-crossing, and stored as a stimulus. After editing, 300 milliseconds of silence was added at the beginning and ending of the selected portion for each stimulus.

In total, 440 stimuli were prepared for the perception experiment. Additional stimuli were prepared for the practice sections. The stimuli for practice were five vowel (/i/) stimuli, ten phrase stimuli (five from English talkers and five from Japanese), and six sentence stimuli (three from English talkers and three from Japanese). They were taken from the talkers different from the sixty talkers in the experiment.

Procedure

After they read and signed the consent form, participants answered the questionnaire. The questionnaire included questions such as demographic information, linguistic and residential backgrounds, and health status with respect to speech and hearing. The questionnaire can be found in Appendix D.

The stimuli were blocked by the stimulus type. Also, the phrase and sentence stimuli were blocked by talkers’ language, while the vowel stimuli were not blocked by talkers’ language. Hence, the stimuli were presented to each listener in five blocks. Three stimuli types and talkers’ languages were presented in counterbalanced order, as listed in Table 14. The stimuli in each block were presented to each listener in random order. Three or five practice trials preceded each section.
Table 14. Order of stimulus presentation. The three stimulus types are indicated as Vowel, Phrase, and Sentence. For the phrase and sentence stimuli, the talkers’ language was noted as E for English, J for Japanese in the parenthesis.

<table>
<thead>
<tr>
<th>Order of stimuli presentation</th>
<th>Number of listeners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>Block 2</td>
</tr>
<tr>
<td>Vowel</td>
<td>Phrase (E)</td>
</tr>
<tr>
<td>Vowel</td>
<td>Phrase (J)</td>
</tr>
<tr>
<td>Sentence(E)</td>
<td>Sentence(J)</td>
</tr>
<tr>
<td>Sentence(J)</td>
<td>Sentence(E)</td>
</tr>
<tr>
<td>Phrase (E)</td>
<td>Phrase (J)</td>
</tr>
<tr>
<td>Phrase (J)</td>
<td>Phrase (E)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

The perception experiment was conducted in a quiet room using scripts in MATLAB version 6.5, which enabled each listener to complete the experiment at his/her own pace. Each stimulus was presented at 44.1 kHz sampling frequency\(^3\) with 16 bits per sample through a headphone (Sennheiser HD250-II or SONY MDR-Z600) connected to a computer at a comfortable listening level. Listeners were asked to listen to each stimulus carefully and estimate how old the speaker was with a scale ranging from 1 to 100 years old. Each judgment was made by clicking a ten-key pad displayed on a computer screen. Each listener could replay each stimulus up to three times if s/he wished. A short break was taken after each section was completed. For the phrase sections, a short break was taken every 90 trials. It took about 2-3 hours to complete the whole experiment.

\(^3\) A high sampling rate of 44.1 kHz was employed because it is a high-fidelity standard and used for a conventional audio CD.
Analysis

In order to see the relationships between talkers’ chronological age (CA) and their perceived age (PA), Pearson correlation coefficients were computed in the three contexts: vowel, phrase, and sentence contexts.

In order to examine the effect of language familiarity, data were grouped by language, the talkers’ language and listeners’ language. That is, if the language of talkers matches with the listeners’ language, this is defined as a familiar language group. If a talkers’ language does not match with listeners’ language, this is a less-familiar language group. Pearson correlation coefficients were computed for these familiar and less-familiar groups. If listeners are able to perceive age for talkers of a familiar language more accurately than talkers of a less-familiar language, then there would be a stronger relationship between CA and PA for the familiar language group than the less-familiar language group. Comparisons of each pair of correlation coefficients were conducted by Fisher’s $r$ to $z$ transformation tests.

In order to examine the results in terms of accuracy of age estimation, the difference of age (hereafter, DA) was calculated by subtracting CA from PA. Three-way ANOVAs were performed on DA with a two-level familiarity effect, a three-level context effect, and a three-level age groups of talkers. Due to a large sample size, a more conservative significance level of $p = .01$ was employed.
Results

Relationships between perceived age and chronological age

It was expected that the relationship between CA and PA would become stronger when the listeners listened to speech that had more information in it. Pearson correlation coefficients between CA and PA for all talkers by all listeners in each context were significant [Vowel: \( r = 0.384 \); Phrase: \( r = 0.621 \), Sentence: \( r = 0.764 \)]. These results indicate that the correlations became stronger as the amount of information increased.

In order to see whether there was a difference between each context pair (i.e., vowel-phrase pair, vowel-sentence pair, and phrase-sentence pair), the \( r \) values in each context mentioned above were \( z \)-transformed, and the differences between the two correlation coefficients were computed. The observed values of \( z \) for each of the context pairs were significantly different.

Table 15. Pearson correlations between PA and CA in terms of language familiarity.

<table>
<thead>
<tr>
<th>Language familiarity</th>
<th>Context</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vowel</td>
<td>Phrase</td>
<td>Sentence</td>
</tr>
<tr>
<td>Familiar language</td>
<td>0.421</td>
<td>0.709</td>
<td>0.847</td>
</tr>
<tr>
<td>Less-familiar language</td>
<td>0.347</td>
<td>0.534</td>
<td>0.687</td>
</tr>
<tr>
<td>All</td>
<td>0.384</td>
<td>0.621</td>
<td>0.764</td>
</tr>
</tbody>
</table>

All correlations are significant, \( p < 0.01 \).

Pearson correlation coefficients were also computed separately for familiar and less-familiar language data in order to see the effect of language familiarity factor on the perceived age. Table 15 is a summary of correlation coefficients with respect to language familiarity. Within each context, \( r \)-values were larger for a familiar language than a less-familiar language. Fisher's \( r \) to \( z \) transformation tests revealed that there were
significant differences between the $r$-value for familiar language data and the $r$-value for less-familiar language data in the phrase context [$z = 7.76, p < .001$] and sentence context [$z = 10.81, p < .001$], but not in the vowel context [$z = 2.33$].

Table 16. Pearson correlation coefficients between talkers’ chronological age and perceived age in Vowel, Phrase, and Sentence contexts.

<table>
<thead>
<tr>
<th>Listeners</th>
<th>Talkers</th>
<th>Vowel</th>
<th>Phrase</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>English</td>
<td>0.332</td>
<td>0.640</td>
<td>0.809</td>
</tr>
<tr>
<td></td>
<td>Japanese</td>
<td>0.266</td>
<td>0.507</td>
<td>0.668</td>
</tr>
<tr>
<td>Japanese</td>
<td>English</td>
<td>0.427</td>
<td>0.590</td>
<td>0.704</td>
</tr>
<tr>
<td></td>
<td>Japanese</td>
<td>0.506</td>
<td>0.793</td>
<td>0.890</td>
</tr>
<tr>
<td>English</td>
<td>English &amp; Japanese</td>
<td>0.298</td>
<td>0.546</td>
<td>0.736</td>
</tr>
<tr>
<td>Japanese</td>
<td>English &amp; Japanese</td>
<td>0.467</td>
<td>0.691</td>
<td>0.792</td>
</tr>
<tr>
<td>English &amp; Japanese</td>
<td>English</td>
<td>0.380</td>
<td>0.612</td>
<td>0.750</td>
</tr>
<tr>
<td>English &amp; Japanese</td>
<td>Japanese</td>
<td>0.388</td>
<td>0.654</td>
<td>0.780</td>
</tr>
</tbody>
</table>

All correlations are significant, $p < 0.01$.

Pearson correlation coefficients were also computed separately for data grouped by listener’s language and talker’s language to examine the specific language factor. Table 16 summarizes all those coefficients. Pearson correlation coefficients were also computed for each listener group for all talkers (the fifth and the sixth rows of coefficients). The last two rows display the coefficients for the English talkers and the Japanese talkers by all the listeners, respectively. Surprisingly, in every context, higher coefficient values were obtained for Japanese listeners than for English listeners as long as the language familiarity is same (see the first vs. fourth rows for familiar language data, the second and the third rows for less-familiar language data, and the fifth and sixth rows for comparison of the two listener languages in Table 16). On the other hand, the language of talkers did not cause significant differences in the correlations in each
context (see the last two rows in Table 16). The reason for the difference between listener groups will be discussed further below.

Accuracy

Differences between perceived age and chronological age (DA) were analyzed in order to examine how talker’s age group and language familiarity would influence the accuracy level of age estimation. If the listeners estimated exactly right, the values of DA would be zero. Therefore, non-zero values of DA indicate error in age estimation. Negative DA values indicate that listeners underestimated the age of talkers, whereas positive DA values indicate listeners’ overestimation. If the age of talkers does not affect the listener’s ability to estimate the age of talkers, levels of age estimation accuracy would be same for all age groups of talkers. If the listeners can estimate the age of talkers more accurately for talkers in a familiar language than talkers in a less-familiar language, the DA values for familiar language talkers would be closer to zero when compared the DA for less-familiar language talkers. However, it was expected that the effect of language familiarity would not appear in the vowel context.
Figure 1. Age estimation errors (=DA) for each age group by familiar vs. less-familiar listener groups in the vowel context.

Figure 2. Age estimation errors (=DA) for each age group by familiar vs. less-familiar listener groups in the phrase context.
Figure 3. Age estimation errors (=DA) for each age group by familiar vs. less-familiar listener groups in the sentence context.

Figure 1, Figure 2, and Figure 3 plot mean values of DA (= PA − CA) for three age groups of talkers in a familiar language vs. a less-familiar language in each of the three contexts. Regardless of language familiarity, the mean DA values are close to zero for the young talkers, and decline for middle-aged to elderly talkers in the three contexts. It is obvious that the listeners underestimated the age of older talkers. The sizes of estimation errors for both older groups were strikingly large, especially for the elderly talkers. On the other hand, age estimation for young talkers was fairly accurate in both listener groups, although the listeners estimated the age of young talkers as slightly older than the actual age in both vowel and phrase contexts. Notice that the directionality of errors for young talkers was different from the ones for the other two age groups in the vowel and phrase contexts (Figure 1 and Figure 2). Similar to the results found in the
previous studies on age perception, overestimation occurred for young talkers, while the ages of the elderly and middle-age talkers were underestimated. This is the expected error directions, but the magnitude of errors was quite large for elderly talkers. The large amount of underestimation for the elderly talkers suggests that the age range employed by the listeners was narrower than the actual age range of talkers in this study. Hence, although the results of high correlations between chronological age and perceived age suggest that the listeners can perceive a whole age distribution fairly accurately, the perceived age for older talkers was shifted downward because the listeners calibrated the age range inaccurately.

Table 17. Analysis of Variance for Accuracy.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context (C)</td>
<td>2</td>
<td>161.46**</td>
<td></td>
</tr>
<tr>
<td>Age groups (A)</td>
<td>2</td>
<td>2257.99**</td>
<td></td>
</tr>
<tr>
<td>Language familiarity (F)</td>
<td>1</td>
<td>11.96*</td>
<td></td>
</tr>
<tr>
<td>C * A</td>
<td>4</td>
<td>90.89**</td>
<td></td>
</tr>
<tr>
<td>C * F</td>
<td>2</td>
<td>35.02**</td>
<td></td>
</tr>
<tr>
<td>A * F</td>
<td>2</td>
<td>53.65**</td>
<td></td>
</tr>
<tr>
<td>C * A * F</td>
<td>4</td>
<td>5.14**</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>8622</td>
<td>(238.63)</td>
<td></td>
</tr>
</tbody>
</table>

Note. The value enclosed in parentheses represent mean square error. *p< .01. **p< .001.

Table 17 includes the results of three-way ANOVA on DA with language familiarity, context, and age group as independent variables. All main and interaction effects were significant. The posthoc tests of Tukey HSD were conducted for pairwise comparisons of context and age group. All pairs of age group and context were significant.

The results of the three-way ANOVA suggests that accuracy of age perception decreased as the age of talkers increased, \( F(2, 8622) = 2257.99, p < .0001 \). Means (with
standard deviations in parentheses) of DA for young, middle-aged, and elderly talkers were 2.81(13.28), −13.98(16.05), and −24.29(18.84), respectively.

There was a significant effect of context on age perception, \(F(2, 8622) = 161.47, p < .0001\). Accuracy of age perception increased from vowel to phrase, and vowel to sentence context [Vowel: \(M(SD) = −15.75\); Phrase: \(M(SD) = −8.51\); Sentence: \(M(SD) = −11.20\)]. Accuracy of age perception was best in the phrase context. This is due to the results for the less-familiar language data, and will be discussed more later in this chapter. Also, age perception was more accurate for talkers of a familiar language than for talkers of a less-familiar language \((F(1, 8622) = 11.96, p < .001)\).

There was also a significant interaction effect between context and language familiarity \((F(2, 8622) = 35.02, p < .0001)\). For talkers of a familiar language, accuracy increased in the order of vowel, phrase, and sentence contexts \((M(SD) = −15.54(22.95), −9.43(16.63), \text{and} −8.77(12.14), \text{respectively})\). On the other hand, the best listeners’ performance was found in the phrase context for the talkers of less-familiar language [phrase: \(M = −7.60, SD = 20.93\); sentence: \(M = −13.63, SD = 16.74\)].

Regardless of language familiarity, the worst age estimations were found for the vowel context [Familiar language: \(M = −15.54, SD = 22.95\); Less-familiar language: \(M = −15.96, SD = 24.31\)]. Mean values of DA between familiar and less-familiar languages were less than two years for each age group in the vowel context (see Figure 1). This small difference was shown to be non-significant in two-way ANOVA conducted on DA with age group and language familiarity for only the vowel context \((F(1, 2874) = .40, p = .53)\).
In order to examine the effect of the listener language on estimation accuracy, further analyses were conducted. The analyses were performed only for the phrase and sentence contexts because the language familiarity did not affect the estimation accuracy in the vowel context (See Figure 1). Three separate three-way ANOVAs were performed on the DA data for each age group with two levels of context effect (phrase and sentence), two levels of familiarity effect, and two levels of listener language effect as the independent variables. The results of the ANOVA are summarized in Table 18. The main effect of language familiarity was significant for the three age groups. The main effect of listener language was significant for young and elderly talkers, but not for middle-aged talkers. The context effect was significant only for the young talkers. Each of the age groups of talkers showed significant two-way and three-way interactions. In order to see the interactions of three terms, mean DA values of each listener group for each talker groups in each context are plotted in Figure 4 through Figure 6.
Table 18. ANOVA results for main effects and interaction effects of context, language familiarity, and listener language on age estimation accuracy for each age group of talkers.

<table>
<thead>
<tr>
<th>Source</th>
<th>Young</th>
<th>Middle-aged</th>
<th>Elderly</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>F</td>
<td>df</td>
</tr>
<tr>
<td>Context (C)</td>
<td>1</td>
<td>126.74 **</td>
<td>1</td>
</tr>
<tr>
<td>Language Familiarity (F)</td>
<td>1</td>
<td>74.10 **</td>
<td>1</td>
</tr>
<tr>
<td>Listener language (L)</td>
<td>1</td>
<td>90.96 **</td>
<td>1</td>
</tr>
<tr>
<td>C * F</td>
<td>1</td>
<td>17.28</td>
<td>1</td>
</tr>
<tr>
<td>C * L</td>
<td>1</td>
<td>1.70 **</td>
<td>1</td>
</tr>
<tr>
<td>F * L</td>
<td>1</td>
<td>13.62 **</td>
<td>1</td>
</tr>
<tr>
<td>C * F * L</td>
<td>1</td>
<td>15.80 **</td>
<td>1</td>
</tr>
<tr>
<td>Error</td>
<td>1912</td>
<td>(126.54)</td>
<td>1912</td>
</tr>
</tbody>
</table>

Note. The value enclosed in parentheses represent mean square error.
* p < .01. ** p < .001.

Figure 4. Accuracy of age estimation in terms of listener language, language familiarity for young talkers in the phrase and sentence contexts.
For the phrase stimuli of a familiar language, Japanese listeners were more accurate than English listeners in estimating the age of talkers (where DA=0 is most accurate). Japanese listeners also showed more accurate age estimation for young talkers of less-familiar language (i.e. English talkers) in the phrase context. However, when estimating the age of middle-age and
elderly talkers of a less-familiar language, Japanese listeners were less accurate than English listeners. In the sentence context, when the listeners estimated talkers of the less-familiar language, there was no clear difference between Japanese listeners and English listeners for all the three talker groups. For talkers of a familiar language in sentence context, Japanese listeners were more accurate than English listeners in estimating the age of middle-age and elderly talkers, whereas English listeners showed a slightly more accurate estimation for young talkers than Japanese listeners. The latter case could be partially explained by a ceiling performance for young talkers in the sentence context.

![Familiar Language data](image)

Figure 7. Average magnitude of age estimation errors (|DA|) in years for the talkers of a familiar language in each context by each listener group.
Figure 8. Average magnitude of age estimation errors (|=DA|) in years for all talkers of a less-familiar language in each context by each listener group.

Figure 4 through Figure 6 also show that Japanese listeners overall tend to be more accurate than English listeners in estimating the age of talkers. In order to see the difference in age estimation accuracy between English and Japanese listeners clearly, average age estimation error for all talkers in each context are shown separately for each listener group in Figure 7 and Figure 8. Significant differences between listener groups were found for the familiar language data in phrase and sentence contexts and for the less-familiar language data in the phrase context. Significant results for the less-familiar language data would be related to the results that Japanese showed large estimation errors for English middle-aged and elderly talkers, which were clear exceptions to the tendency of more accurate estimation by Japanese. This suggests that the phrase and sentence stimuli by English talkers were perceived as young by Japanese listeners. On the other hand, accuracy for the English listeners was better than the Japanese listeners for
older talkers in phrase context. However, considering the underestimation tendency observed in
the English listeners, more accurate estimation actually means that the English listeners tended to
estimate the Japanese talkers as relatively old in their age scale. This might be due to the
difference in the dental condition between Japanese and English talkers. Seven of ten Japanese
elderly talkers wore either partial or complete denture and one Japanese elderly man (JMO05)
had only several teeth but did not wear denture, while the elderly talkers of English reported that
their dental conditions were fairly good (see Appendix A1 and A2). Both denture and missing
teeth could introduce some acoustic characteristics in speech that the listeners can use as
perceptual cues to identify the age of a talker.

In addition, this might be due to a different pronunciation of the fricative /s/ between
English and Japanese. In Japanese, /si/ is pronounced with an alveolo-palatal (or laminodorsoro-
alveolopalatal) fricative [ɕ]. For the English listeners, fricatives in the Japanese phrase stimuli
could sound ‘funny’ or ‘imprecise’. Discrepancy from the norm might be perceived as a mild
degree of speech disorder. According to Ryan and Burk (1974), “individuals judged to be older
may be speakers who would fall at the mild end of a “dysarthria continuum””. Hence, deviant
pronunciation from English by Japanese talkers might increase the perceived age for Japanese
phrases by English listeners. On the other hand, Japanese listeners might notice some acoustic
differences in the English phrase stimuli, but it would be less likely that Japanese listeners
perceived the pronunciations of /si/ by the English talkers [si] as something so deviant from
Japanese pronunciations, because an alveolo-palatal fricative is an allophonic variant of /s/ in
Japanese, and the phoneme /s/ is also pronounced as [s], rather similar to an English [s].
Discussion

The results of the current study found that the young listeners aged 18-28 years exhibited moderate to high correlations between talker’s age and perceived age confirming results found in the previous literature. In order to directly compare the current results with the previous results, Pearson correlation coefficients were computed between CA and mean PA of all listeners in each context because the correlations in the previous studies were obtained from mean perceived age not from perceived age. The correlation results with mean PA are similar to the correlations with PA. In the present study, correlations in the three contexts were significant [Vowel: \( r = 0.626 \); Phrase: \( r = 0.817 \), Sentence: \( r = 0.871 \)]. Especially, the \( r \) value in the sentence context, which contained the most similar stimuli to those used in previous studies, is similar to the values reported in those studies (See Table 10). The slightly higher correlation values in this study compared to some of the previous studies could be attributable to methodological differences. There was not much difference between the current study and the previous studies in terms of the stimuli length. The length of our single sentence was shorter than some of the previous studies that employed the entire “Rainbow Passage” as their stimuli (Baker, 1981; Kukol, 1979). Nevertheless, similar correlation values were obtained in the current study, where listeners heard much shorter speech material. This suggests that abundant information to indicate age of talkers is available in a single sentence. Another notable difference may be the age range of the talkers. Because correlation is especially sensitive to extreme values, a higher correlation value could be related with a wider age range of the talkers. However, unlike some previous studies (e.g. Shipp & Hollien, 1969; Ryan & Burk, 1974), listeners in the current study were not informed of the actual age range of the speakers they listened to. Therefore, the range of perceived age may not automatically expand to the age range corresponding to the talkers. Hence, it is not likely that the
higher correlations were derived from a relatively wide range for talkers’ age employed in this study.

The effect of the amount of information is clear. The result that the correlation coefficients increased from the vowel to phrase to sentence contexts confirmed the prediction that more information helps the perception of age regardless of the listeners’ language, or familiarity with the talker’s language. It suggests that listeners can perceive talkers’ age more accurately as the information contained in the speech increases. It also suggests that voice characteristics available from sustained vowels alone could provide some age information. Comparing the results from the phrase stimuli with ones from the vowels, there is a large improvement of listeners’ performance in the phrase context. This may be due to variation in the length of stimuli. While editing the original utterances in order to create the stimuli, the author noted that the phrase stimuli of the elderly talkers tended to be longer than the phrase stimuli of the younger talkers. Such individual differences in phrase duration could provide listeners critical information associated with talker’s age. Whether or not there is rate information, the phrase stimuli also include dynamic changes, such as changes in the pitch contours and information about articulator movements that occurred within the phrase. Although some studies have showed that mean speaking fundamental frequency was not well correlated with perceived age (e.g. Braun & Rietveld, 1995), it is possible that dynamic changes in fundamental frequency are closely associated with perception of talker’s age.

Our results also found that language familiarity is an important factor for age perception. As expected, language familiarity does not affect age perception when the listeners judged the talker’s age from the sustained vowel alone. On the other hand, the influence of language familiarity was great when the listeners judged the phrase and sentence stimuli. In general, age
perception was more accurate for the familiar language talkers than the less-familiar language talkers. One thing to note is that unlike the phrase and sentence stimuli, the vowel stimuli were not blocked by talker’s native language when they were presented to the listeners. This difference may account for the different results found in the vowel although the listeners told that they could not notice whether the vowel stimuli were produced by native or non-native speakers.

The most accurate age estimation was obtained for the young talkers, while a considerable underestimation was found for the middle-age and elderly talkers. The participants in the current study included only young listeners. Therefore, the result of better judgments for young talkers can be accounted for by a peer-group factor. The frequent and possibly influential social interactions of young listeners with young talkers might increase their ability to judge the age of young talkers. In fact, a lot of listeners said in a post-experiment interview that they think that old people for them are people in their late forties or fifties. This matches with the result that the mean perceived age for the elderly talkers was 58.01 years old. This is far below the corresponding actual average age for the elderly talkers ($M = 82.3$ years). It was expected that listeners exhibit more accurate age judgments for talkers of similar ages to the listeners, or vice versa. Neiman and Applegate (1996) found a similar result of more accurate perception of age for the young talkers by the young listeners and less accurate results by the same young listeners for elderly talkers. Also, Huntley et al. (1987) found that elderly listeners overestimated the age of the young talkers more than the young listeners did. It would be interesting to see whether there would be an improvement in the listeners’ overall performance if listeners were matched to the same age brackets of the middle-age or elderly talkers.

One difference between the two listener groups was the ratio of listener sex. It happened that Japanese listeners were all female. On the other hand, English listeners consisted of both
men and women. This difference in male-female ratio of the listeners may be a cause of the difference between English and Japanese listeners. To check this possibility, correlations for each context were recalculated with the data from just the female listeners and are listed in Table 19. Excluding male listeners did not improve the correlation values for English data. Therefore, it is unlikely that the male-female ratio difference between Japanese and English listeners is the reason for higher correlations obtained from Japanese listeners.

Table 19. Pearson correlation coefficients between CA and PA in each context based on the data from female listeners.

<table>
<thead>
<tr>
<th>Listeners</th>
<th>Talkers</th>
<th>Vowel</th>
<th>Phrase</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>English (female)</td>
<td>English &amp; Japanese</td>
<td>0.262</td>
<td>0.550</td>
<td>0.726</td>
</tr>
<tr>
<td>Japanese</td>
<td>English &amp; Japanese</td>
<td>0.467</td>
<td>0.691</td>
<td>0.792</td>
</tr>
</tbody>
</table>

Another possible account for the result that Japanese listeners are better than English listeners at judging age of talkers is differences in the residential environment between the two listener groups. Out of 24 Japanese listeners, fourteen listeners lived with their grandparent(s). On the other hand, none of the English listeners lived with a grandparent. Of course, lack of residential experience with grandparents does not mean that the English listeners had no social interaction with elderly people. However, if we take the residential environments as a rough index for the amount of social interactions with elderly people, the Japanese listeners had more opportunity to communicatively interact with elderly people than the English listeners. More frequent inter-generational communications may enable the Japanese listeners to use acoustic characteristics which are well-correlated with talker’s age more easily than the English listeners.

Out of five possible outcomes of the accuracy of age estimation described in page 56, results of the current study are similar to the possible outcome (5) in page 56. However, there was a difference in terms of listener’s native language. The outcomes in this study are expressed
in (6) for English listeners and (7) for Japanese listeners. English listeners exhibited exactly the same pattern as in (5). The results of Japanese listeners were similar to (5), but the effect of language familiarity is greater than the context effect. As for the age estimation in the familiar language, Japanese listeners estimated the age of talkers more accurately than English listeners. Factors to determine the ranking order of languages are not clear, but the language effect is tentatively expressed as in (8). Further investigation in different languages would reveal which of the two patterns is more commonly observed across languages.

(6) English listeners:
   Vowel (L=F) < Phrase (L<F) < Sentence (L<F)

(7) Japanese listeners:
   Vowel (L=F) ≤ Phrase (L) < Sentence (L) < Phrase (F) < Sentence (F)

(8) English listeners < Japanese listeners,
   only if they judge the age of speakers in one’s own native language.

Conclusions

This study explored the effect of language familiarity on perception of talker’s age at different levels of linguistic context. Age estimates were more accurate for the talkers in the familiar language than the less-familiar language. The language familiarity effect was found at the phrase and sentence level, but not at the vowel level. Also, age estimation accuracy improved as the amount of information increased. These results suggest that in addition to physiologically determined aspects of voice characteristics, dynamic properties of speech observed in speech productions, even within the length of only three syllables, embody important
properties associated with aging. Furthermore, familiarity of language and possible familiarity with non-peer groups are important factors to influence the perception of talkers’ age. And finally, the result that Japanese listeners were better judges than English listeners suggests that listeners’ sociolinguistic backgrounds affect age perception. The results of familiarity effects suggest that perception of talker’s age is based on listeners’ perceptual experiences including language-specific factors associated with age. Therefore, we need to be cautious in assuming that acoustic characteristics associated with judging aging speech in one language apply to another one. Furthermore, perceptual judgments about speech in a foreign language, or by any speaker of an unfamiliar linguistic background, should be viewed with considerable caution.
CHAPTER 4

4. The effects of talker’s sex

Introduction

Previous literature on age perception assumes that perceptual cues for talker’s age are based mostly on physiologically determined factors. There are various physiological changes that occur with advancing aging. However, the age-related changes in the voicing mechanism do not always affect males and females in the same way. Generally, more substantial changes occur in males than females. For example, calcification of laryngeal cartridges and stiffening and thinning of vocal fold tissues and fibers are extensively observed in males (Linville, 2001). Also, these age-related changes in laryngeal structures begin earlier in the life for men than women (Linville, 2001).

In fact, the life expectancy of women is longer than men in general. The sex differences in mortality through one’s life and hormonal differences between men and women suggest that women have the biological advantage (Desjardins, 2004). The incidences of various biological markers of aging are higher in men than women in industrialized countries (Goldman et al., 2004; Price, Fowkes, & Gerald, 1997). Elderly speakers in poor physiological condition were perceived as older than the age-matched speakers in good health condition (Ramig, 1986). Hence, if sex related differences in speech are compared using the speaker’s chronological age, especially among the elderly
population, it is often the case that the aging process is less advanced in older women compared to chronological age-matched men\(^4\).

Sex differences in speech have been also investigated in the field of sociolinguistics. Sociolinguistic studies have documented speech variation related to the speaker’s sex. Besides choice of vocabulary, gender differences in language have been observed in phenomena at lower linguistic levels such as in differences in vowel quality, or different degrees of occurrences of phonemic deletion or assimilation, and in higher level phenomena such as frequency of interruption in conversation involving both sexes. Sex differences found in historic sound change are particularly interesting phenomena. Women use the advanced forms of sound change more often than men in most cases (e.g. Labov, 1990, 2001). That is, a series of sound changes can be easier to document by examining female speech. Men are more reluctant to accept the advanced forms, but rather prefer to use the more conservative forms, even though they may be socially stigmatized (Labov, 1990, 2001). Based on sex differences observed in ongoing sound change, Labov (1990, 2001) proposed that women lead linguistic change because of the active and frequent use of more progressive linguistic forms by women. Hence, because women accept contemporary forms, while men remain using the traditional forms, sociolinguistic variation is greater in women than men.

Another noteworthy sociolinguistic phenomenon is sex difference in average pitch level. Japanese females are known to employ higher parts of their pitch range, compared with women in other countries such as the United States (Loveday, 1981; Yamazawa & Hollien, 1992; Ohara, 1992) or the Netherlands (van Bezooijen, 1995).

\(^4\) This may not be true for people at age 85 or older (Perls, 1995).
This tendency is usually explained from a sociolinguistic point of view. That is, Japanese females subconsciously respond to social expectations or preferences because higher pitch is associated with psychological images of smallness, obedience, and powerlessness (Ohala, 1994), all of which are strongly correlated with an attribute of the ideal Japanese woman (van Bezooijen, 1995). Loveday (1981) compared the average speaking fundamental frequency of Japanese and English men and women. He found that Japanese females tend to speak with a higher fundamental frequency than English females. Loveday proposed that Japanese females employ extremely high pitch in order to express politeness, because the socio-cultural expectation to be polite is higher for females than males in Japan. Ohara (1992) conducted a similar study on bilinguals in Japanese and English, and found that Japanese females speak with higher pitch when they speak in Japanese than when they speak in English. Japanese men, on the other hand, did not show such differences in the two speaking conditions. Her results support Loveday’s interpretations, because social constraints on gender roles should loosen in the English speaking condition when Japanese women are not required to cope with Japanese expectations.

In spite of the fact that there are a number of sex differences related directly to biological foundations, as well as socio-cultural backgrounds, the majority of research on age perception has focused on male speech. There are only a few studies that investigated age perception for both male and female speech. Even among them, the results are not conclusive. Kukol (1979) investigated the relation between speaker’s sex and age perception. He employed a method used in earlier studies of age perception. Twenty-two female listeners directly estimated the age of 13 male talkers and 17 female talkers
aged 40-80 years. He computed a correlation coefficient between perceived age and chronological age from each listener. The median correlation coefficient between perceived age and chronological age for male speakers was $r = 0.80$, whereas the median correlation value for the female speakers was $r = 0.63$. The correlation for male talkers is similar to that observed in other studies. However, because greater variability was found for female speech than male speech, the results suggest that it was more difficult for the listeners to estimate the age of females. He proposed that, either fewer or finer perceptual cues may be needed for listeners to be able to estimate the age of women than men.

However, in a similar study conducted by Baker (1981a), opposite results were reported. Twenty African-American listeners and twenty Caucasian listeners estimated the age of 82 African-American men and 90 African-American women. Baker did not report the correlation values for male and female talkers, because the main focus of her study was the relationships between ethnicity and perceived age. However, because she reported the perceived age of individual speakers (cf. Baker, 1981, pp. 93-104), the correlation value between perceived age and chronological age can be computed separately for male and female speakers for each listener group. The correlation coefficients are summarized in Table 20. Whether the listeners were Caucasian or African-American, correlations were higher for female speakers than male speakers in Baker (1981). Similarly, Neiman and Applegate (1990) found that the correlation between perceived age and chronological age was stronger for female talkers than for male talkers.
Table 20. Correlation coefficients for male and female speech obtained from various studies.

<table>
<thead>
<tr>
<th>Source</th>
<th>Correlations (r-values)</th>
<th>Age range of speakers</th>
<th>Speaker sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kukol (1979)*</td>
<td>0.80</td>
<td>40-80</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>0.63</td>
<td>40-80</td>
<td>Female</td>
</tr>
<tr>
<td>Baker (1981)**</td>
<td>0.64 (Caucasian listeners)</td>
<td>40-81</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>0.70 (Caucasian listeners)</td>
<td>40-95</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>0.64 (African-American listeners)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.73 (African-American listeners)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neiman &amp; Applegate (1990)</td>
<td>0.88</td>
<td>20-75</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>20-75</td>
<td>Female</td>
</tr>
</tbody>
</table>

* Median of correlations from each listener.
** Correlation values were calculated based on the data by the author.

Examining the distribution of sexes in the population, the ratio of men and women is about same in both Japan and the United States. However, given that women have a longer life expectancy than men, the female to male sex ratio increases with age. The female to male sex ratio starts to increase after age 65. Among the population aged 75-84 years in the United States, there are 153 women for every 100 men (U.S. Census, 2000). Ramig and Ringel (1983) and Ramig (1983a) found strong correlations between a speakers’ physiological condition and several acoustic characteristics of speech. Considering that women have a longer average life expectancy than men, it appears that elderly women in general are in better health than men of the same chronological age. Logically, when listeners estimate the ages of men and women of the same chronological age, the perceived age for females should be younger than that for males. However, if listeners normalize the age distributions of males and females, or they can differentiate the acoustic characteristics for male and female, then a talker sex should not influence a listener’s age estimation.
Overall, the influence of talker’s sex on the perception of age has not been studied from the sociolinguistic point of view. Specifically research is needed to determine whether sociolinguistic factors such as age, gender, and native language, influence age perception.

**Purpose**

This chapter presents an evaluation of the sociolinguistic component in age perception using the same data described in the previous chapter. The data were separated in terms of the language of the talkers, and a separate analysis was conducted for each data set. Differences in listeners’ performance for male and female talkers were hypothesized based on three accounts described below. These predictions for listeners’ performance are summarized in Table 21. Only predictions for the responses to the English talkers are shown here, because similar predictions for the responses to the Japanese talkers are expected.

<table>
<thead>
<tr>
<th>Talker language</th>
<th>Listener language</th>
<th>Biological based account</th>
<th>Demographic based account</th>
<th>Sociolinguistic account</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>English</td>
<td>M&gt;F</td>
<td>M&lt;F or M=F</td>
<td>M&lt;F</td>
</tr>
<tr>
<td>English</td>
<td>Japanese</td>
<td>Same as above.</td>
<td>Same as above.</td>
<td>M=F</td>
</tr>
</tbody>
</table>

In a biologically based account, the age of talkers is judged based on the acoustic characteristics due to the physiological changes by aging. The perception of age should be better for male speech than female speech because there is a greater variability due to aging found in male voices for many acoustic characteristics (Linville, 2001). If the
same acoustic characteristics are used by listeners to estimate the age of men and the age of women, it is expected that male speech would be perceived as older than female speech. And more importantly, the same female-male relationship should be found in both language groups.

On the other hand, if the listeners’ age estimation is influenced by demographic (or population) sex differences in the male-female ratio, better performance for female speech is expected as a consequence of decreasing numbers in the men. That is, the listeners would have more communicative opportunities with a wider age range of female talkers than male talkers such that they could store more detailed speech information about female speech. Because the sex ratios in the United States and Japan are similar, a similar pattern of sex differences in both language groups is expected. However, if the listeners somehow normalize the sex differences in the population distributions, no differences related with speakers’ sex will be observed.

Finally, if the sociolinguistic factors affect the listeners’ perception of the talker’s age, better listener performance for female speech than male speech is predicted. Young to middle-age females adopt more advanced forms of the active linguistic change, whereas the males in the same community tend to use the stigmatized form across generations. Thus, more sociolinguistic characteristics that could indicate a talker’s age are expected to be found in the female speech and therefore, it might be easy for listeners to determine the age of female talkers. However, such linguistic knowledge is not accessible to the members outside of the language community. Hence, the sex differences would not be observed when the listeners judge the age of talkers who are from a different language community.
Methods

The data to be analyzed in this chapter were the same data described in the previous chapter, but subdivided by talker’s sex. Analyses were conducted only on the data from the sentence stimuli based on three reasons. First, because the sentence stimuli were the most similar material to those used in the previous experiments, the results can be easily compared with previous results. Secondly, the best listeners’ performance was obtained from the sentence stimuli. And, finally, it is considered that if there are any, sociolinguistic factors would be most available from the sentence stimuli. Therefore, the data to be analyzed here includes perceived age results obtained for sixty sentence stimuli. The stimuli consisted of five male and five female speakers from each of the three age groups for each language group. Table 22 summarizes mean and range of age for each subgroup of talkers in each language group.

<table>
<thead>
<tr>
<th>Age groups</th>
<th>Talker’s language</th>
<th>English</th>
<th></th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Talker’s sex</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Young</td>
<td>N=5</td>
<td>25.4(24-28)</td>
<td>26.2(24-30)</td>
<td>N=5</td>
</tr>
<tr>
<td></td>
<td>N=5</td>
<td>57.8(56-60)</td>
<td>56.2(54-59)</td>
<td>N=5</td>
</tr>
<tr>
<td>Middle</td>
<td>N=5</td>
<td>82.2(80-83)</td>
<td>82.0(80-86)</td>
<td>N=5</td>
</tr>
<tr>
<td>Elderly</td>
<td>N=5</td>
<td>82.2(80-83)</td>
<td>82.0(80-86)</td>
<td>N=5</td>
</tr>
</tbody>
</table>
Listeners

Twenty-four native speakers of American English (15 female and 9 male) and twenty-four native speakers of Japanese (24 female) served as listeners in the study. They are the same subjects described in the previous chapter.

Analysis

Listeners’ responses for English talkers and their responses for Japanese talkers were analyzed separately. Pearson correlation coefficients between talkers’ chronological age (CA) and their perceived age (PA) were computed for female and male stimuli. Accuracy of age estimation was measured by a difference between CA and PA (DA= PA−CA), and used as a dependent variable for the ANOVAs.

Results

Perceived age for English talkers

Perceived age

Table 23 summarizes the Pearson correlation coefficients between talkers’ chronological age (CA) and their perceived age (PA) by each listener group with respect to the sentences spoken by male and female speakers. When English listeners judged the age of English talkers, correlation was stronger in female speech (r = 0.86) than male speech (r = 0.76). A Fisher's r to z transformation test revealed that the difference between the two correlations was significant (z= 4.15, p < .001). However, when the same sentence stimuli were judged by the Japanese listeners, the difference found between male and female speech disappeared (r = 0.72 for male talkers and r = 0.74 for
female talkers). Fisher’s transformed $z$ did not reach a significant level of difference between the two correlations ($z=-0.48, p = .63$).

English listeners showed a stronger correlation result for English female speakers ($r = .86$) than Japanese listeners ($r = .74$). Comparing the Fisher’s $r$-to-$z$ transformed correlation coefficients showed a significant difference between the two correlations ($z=4.72, p < .001$). This suggests that the variability in age estimation for female talkers was higher among Japanese listeners than English listeners. However, correlations for English male speech were not significantly different in the two listener groups ($z=0.89, p = .29$).

**Table 23. Correlation between CA and PA for English talkers (n=360). All correlations are significant at $p<.0001$.**

<table>
<thead>
<tr>
<th>Talkers</th>
<th>English listeners</th>
<th>Japanese listeners</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Female</td>
<td>0.861</td>
<td>0.737</td>
</tr>
<tr>
<td>English Male</td>
<td>0.756</td>
<td>0.720</td>
</tr>
</tbody>
</table>

Correlation results suggested that variability in PA was greater for male talkers than female talkers. This is supported by the difference in standard deviation between male and female talkers. **Table 24** displays means and standard deviations of PA by English listeners for English male and female talkers. English male talkers exhibited higher standard deviations than English female talkers.
Table 24. Means and standard deviations of PA for English male and female talkers (young, middle-aged, and elderly) by English listeners (n=360).

<table>
<thead>
<tr>
<th></th>
<th>English female talkers</th>
<th>English male talkers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Young</td>
<td>25.7</td>
<td>7.9</td>
</tr>
<tr>
<td>Middle-age</td>
<td>44.3</td>
<td>9.5</td>
</tr>
<tr>
<td>Elderly</td>
<td>64.3</td>
<td>10.3</td>
</tr>
</tbody>
</table>

The difference in age estimation between male and female talkers by English listeners could be related to listener’s sex. Because English perception results were obtained from both male (n=9) and female (n=15) English listeners, the effect of listener’s sex on age perception was examined by sorting the data by listener’s sex.

Table 25 shows correlations between CA and PA for English male and female speech age-estimated by male and female listeners. There was still a clear difference in terms of talker’s sex. The correlation obtained for English female talkers ($r=0.861$) and the correlation obtained for English male talkers ($r=0.756$) were significantly different regardless of listener’s sex in Fisher’s $r$-to-$z$ transform tests ($p<.01$). However, no significant differences were found between English male and female listeners (correlations for all English talkers were $r=0.812$ for male listeners and $r=0.809$ for female listeners). Correlations obtained from English female and male listeners (see Table 25 for the $r$-values) were not significantly different for English female talkers ($p=.62$), nor were they for English male talkers ($p=.88$). These results suggest that female listeners do not show an advantage of being female when estimating the age of female talkers, and correspondingly male listeners do not have a disadvantage of being male when estimating the age of opposite-sex talkers.
Table 25. Correlation between CA and PA for English male and female talkers by English male and female listeners. Sample numbers for correlation coefficients calculated from each female listener were 225 and sample numbers for correlation coefficients from each male listener were 135.

<table>
<thead>
<tr>
<th>Talkers</th>
<th>English female listener</th>
<th>English male listener</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Female</td>
<td>0.856</td>
<td>0.870</td>
</tr>
<tr>
<td>English Male</td>
<td>0.763</td>
<td>0.755</td>
</tr>
</tbody>
</table>

Accuracy

In order to examine the differences between English male and female speech in terms of accuracy of age estimation, the DA values (=PA−CA) for male and female speech were compared. A negative value for DA means age was underestimated. Figure 9 and Figure 10 shows the accuracy of age estimation for English talkers in terms of talkers’ sex and age groups. Figure 9 displays DA from English listeners, and Figure 10 displays the results from Japanese listeners. First, in both listener groups, age was underestimated for both men and women in the middle-age and elderly English groups. However, for the speech of young talkers, a slight overestimation was found for male speech from both English and Japanese listeners. However, as seen on Figure 9, English listeners exhibited only a small difference between male and female speech, whereas Japanese listeners perceived the age of female talkers as much younger than male talkers across all age groups (see Figure 10).
Figure 9. Difference between perceived age and chronological age (DA) in terms of talkers’ sex and age groups for the English sentence stimuli by English listeners. DA for male talkers is indicated by squares and for female talkers by triangles.
Age estimation for English sentences by Japanese listeners

Talkers:  
- male  
- female

Figure 10. Difference between perceived age and chronological age (DA) in terms of talkers’ sex and age groups for the English sentence stimuli by Japanese listeners. DA for male talkers is indicated by squares and for female talkers by triangles.

Separate two-way ANOVAs for each listener group using DA as a dependent variable and talkers’ sex and age groups as independent factors revealed that the difference in DA between male and female speech was highly significant in Japanese listeners’ performance ($F(1, 714) = 113.45, p < .0001$), but marginally significant in the results from English listeners ($F(1, 714) = 5.50, p = .02$). The effect of age group was significant in both listener groups [English listeners: $F(2, 714) = 244.48, p < .0001$, Japanese listeners: $F(2, 714) = 273.85, p < .0001$]. The interaction effect was not significant for English listeners ($p = .05$), but was significant for Japanese listeners ($F(2, 714) = 7.82, p < .0001$). This can be accounted for by a small difference in DA for middle-age female and elderly female (Means of DA were -25.4 years and -27.4 years,
respectively). Sheffe’s post-hoc analysis found no significant difference in DA for middle-age and elderly female talkers in terms of the estimation by Japanese listeners. Results suggest that Japanese listeners perceived English female talkers as particularly young.

**Perceived age for Japanese talkers**

**Perceived age**

Table 26 presents r-values for each listener group in terms of Japanese male and female talkers in the sentence context. A pair of the correlation coefficients was compared using Fisher’s r-to-z transformation. As seen in Table 26, the correlation between CA and PA was stronger for female talkers than male talkers (p< .01) when the Japanese listeners listened to the Japanese speakers, whereas no significant difference for talker sex was found when English listeners judged the age of Japanese talkers (p= .42). These results are exactly the same in the results obtained from English talkers. Table 27 displays means and standard deviations of PA for Japanese talkers. As observed in English results above, standard deviations were also smaller for female talkers than male talkers in all three age groups of Japanese talkers. Hence, Japanese listeners also exhibited less variability in PA for female talkers than male talkers when they judged the age of talkers in their own language.
Table 26. Correlations between CA and PA for Japanese talkers (n=360). All correlations were significant at $p<.0001$.

<table>
<thead>
<tr>
<th>Talkers</th>
<th>English listener</th>
<th>Japanese listener</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese Female</td>
<td>0.686</td>
<td>0.930</td>
</tr>
<tr>
<td>Japanese Male</td>
<td>0.653</td>
<td>0.850</td>
</tr>
</tbody>
</table>

Table 27. Means and standard deviations of PA for Japanese male and female talkers (young, middle-aged, and elderly) by Japanese listeners (n=360).

<table>
<thead>
<tr>
<th></th>
<th>Japanese female talkers</th>
<th>Japanese male talkers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Young</td>
<td>20.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Middle-age</td>
<td>50.7</td>
<td>8.3</td>
</tr>
<tr>
<td>Elderly</td>
<td>70.9</td>
<td>9.9</td>
</tr>
</tbody>
</table>

Accuracy

The analysis of age estimation accuracy for Japanese talkers revealed interesting results. DAs for male and female Japanese talkers in each age group are displayed in Figure 11 and Figure 12. Figure 11 shows the results from a group of English listeners, whereas Figure 12 shows the results from the Japanese listener group. Again, two two-way ANOVAs were performed on DA for Japanese talkers from English listeners and Japanese listeners with talker’s sex and age groups as independent factors.
Figure 11. Difference between perceived age and chronological age (DA) in terms of talkers’ sex and age group for the Japanese sentence stimuli by English listeners. DA for male talkers is indicated by squares and for female talkers by triangles.
Figure 12. Difference between perceived age and chronological age (DA) in terms of talkers’ sex and age group for the Japanese sentence stimuli by Japanese listeners. DA for male talkers is indicated by squares and for female talkers by triangles.

As shown in these figures, there was no significant sex-related difference for Japanese talkers in either listener groups. Indeed, the main effect of talker sex was not significant in either listener group, $p = .25$ for English listeners, and $p = .91$ for Japanese listeners. In both listener groups the effect of age group was significant ($F(2, 714)=1.83$, $p< .0001$ for English listeners, and $F(2, 714)= 2.90$, $p< .0001$ for Japanese listeners), but the interaction was not significant ($p = .16$ for English listeners, and $p = .06$ for Japanese listeners). These results suggest that whether the talker is male or female, both listeners increased the estimation errors when the talkers were older. However, as seen in Figure 11 and Figure 12, it seems that Japanese listeners exhibit more accurate age estimation than English listeners. There are two clear differences between the two listener groups.
First, estimation errors found for young and middle-age talkers were approximately the same for the Japanese listeners. The difference between mean DA for young talkers and mean DA for middle-age talkers was 0.7 years in the Japanese listener group, but 14.2 years in the English listener group. In fact, Sheffe’s post-hoc analyses found that DAs obtained for middle-age and elderly talkers by Japanese listeners were not significantly different with respect to men and women. Furthermore, Japanese listeners showed smaller estimation errors for the elderly men and women than English listeners, even though the worst performance was found for the elderly group in both listener groups. These results suggest that Japanese listeners were able to estimate the age of Japanese talkers more accurately than English listeners.

**Discussion**

The ability to estimate age of male versus female talkers was examined in terms of correlation and accuracy. Correlation results obtained from each listener’s age estimates for all talkers indicate how strong the relationship of CA and PA is among the listeners. In other words, higher correlation means less variability in listeners’ age estimates, but does not necessary mean high accuracy. At first, the results of correlations and accuracy in this study seem inconsistent with one another. Sex-related differences were obtained in the correlation results, but not in terms of accuracy, except for Japanese age estimates for English female talkers. That is, the exceptional case for English female speech seems to require a different account (presented below), but aside from that, the results of this study suggest that listeners estimate age at about same accuracy whether the talker is male or female, but variability of perceived age was greater for male talkers than female talkers. Less variable age estimation for female talkers suggests that listeners
had the ability to make more fine-grained age distinctions for females than males. Of course, correlation results could be influenced by the actual age differences between male and female talkers because male and female talkers were not exactly age-matched (see Table 22). However, this does not account for the result that the sex-related correlation difference was not found between English and Japanese listeners.

The question is then why listeners had a tight connection between talker’s age and female speech. Among the three accounts proposed in the introduction, only the sociolinguistic account can fully account for the sex difference in the current study. Overall results indicated that listeners exhibited a better age judgments for female speech. Also, results demonstrated an asymmetrical sex difference between English and Japanese listeners. The sex difference was only found when the listeners and the talkers were from the same language community. The sociolinguistic account expected that the sex differences would be found only when listeners and talkers were from the same language community. On the other hand, neither of the other two accounts is supported by the current results given that the sex differences disappeared in the language mismatched data.

In addition, strong correlations between PA and CA found for females may be related with a female gender role or social expectations toward females. Women’s attractiveness is depicted stereotypically with the age attribute of ‘young’ in the advertisements in many Western countries (Michell & Taylor, 1990; Gilly, 1988) and Japan (Ford, Voli, Honeycutt, & Casey, 1998). Since age is more important factor for women than for men, listeners may pay more attention to the age-related properties in
speech for female speech than for male speech. Hence, the listener’s age estimation was better for female speech than male speech.

One odd finding was that Japanese listeners estimated the age of English females particularly young (see Figure 10). Perhaps this was due to some sociolinguistic convention that is only prevalent in Japanese society. English male speech might be perceived relatively old compared with English female speakers because acoustic characteristics associated with advanced aging for Japanese listeners are more salient in male speech. However, the sex-related difference was only seen when Japanese listeners estimated age for the less familiar language, English. Since Japanese listeners did not exhibit sex-related differences in age perception for their own language, underestimation of the age of English females by Japanese would be derived from a certain acoustic properties in the English female speech. And these cues must be expected to be found in the younger female speech, not in the middle-age female speech.

It seems (from the author’s observation) that Japanese female speakers, especially elderly females, speak in a soft voice. It is known that Japanese learners of English are sometimes misperceived because they do not speak loud enough. This is not simply due to shyness or a reserved attitude common to Japanese speakers. Abdominal respiration is not commonly used in Japanese speech production. Keithly (1999) reported that Japanese ESL speakers use a different respiration pattern when they read sentences in English and when they speak in Japanese. It would be interesting to further investigate acoustic characteristics related with loudness or breath group to determine the relationship between certain indexical properties specific to a language community and linguistic properties associated with the phonological system in the language.
Another noteworthy difference due to the language is the accurate age estimation by Japanese listeners for Japanese talkers. This might be explained by the social attitudes toward the elderly people in Japan. Even though people in both the United States and Japan viewed the elderly or advancing aging in general negatively, elderly people are treated with greater respect in Japan than the United States (Foos & Clark, 2002, p.211). Although a recent report indicates that the concept of filial piety is collapsing in Japan (Yamamoto & Wallhagen, 1997), Japanese elderly parents still co-reside with their own children if possible (J. W. Brown et al., 2002). Similarly, many Japanese elderly still prefer to live with family members than live independently. Living in a retirement home is considered as a failed post-retirement life in Japan (Bethel, 1992), which is completely the opposite expectation of elderly people in the United States who highly appreciate an independent living arrangement. Hence, there might be relatively less social pressure to avoid sounding older in Japan. Also, because of the complicated honorific styles of speech, Japanese must carefully attend to the age of people in the conversations in order to speak appropriately. As a result, it is possible that Japanese listeners acquire more detailed knowledge about age-associated speech characteristics than American English listeners.

Alternatively, the dominant acoustic cues that Japanese listeners depend on may be simply different from the dominant acoustic cues used by English listeners. Hence, Japanese listeners might have an advantage over English listeners if their major perceptual cues to identify the age of talkers are acoustically more salient than the cues used by the English listeners.
Conclusions

Listeners exhibited a higher correlation between perceived age and chronological age for female talkers than male talkers when they estimated the age of talkers who were from the same language community. However, when the same speech was judged by the listeners who belonged to a different language community from the speakers, the sex difference previously found for native listeners disappeared. In addition, the age of English females was underestimated by the Japanese listeners. Also, the age estimation for Japanese talkers by Japanese listeners were very accurate compared with the estimation for English talkers by English listeners. These results strongly suggest that sociolinguistic factors play an important role in the perception of age. Furthermore, the strong correlation between chronological age and perceived age for female speech suggests a finer-grained age distinction for female talkers. This may be due to more diverse sociolinguistic variation and an inseparable connection between age and female gender.
CHAPTER 5

5. The effects of age disguise

Introduction

The previous chapters showed that the current study replicated findings in previous literature that a speaker’s chronological age and perceived age are highly correlated with each other when the listeners judged short phrases or sentences spoken in their native language. This finding was also extended to judgments of age for non-native speakers. However, the listeners exhibited better age estimation for a familiar (native) language compared to a less familiar (foreign) language. These results suggest that while the perception of age is based on physiological changes that any individual may undergo in the course of aging, age perception is also heavily influenced by language specific differences. In other words, listeners store certain speech characteristics as properties of speaker’s age, and some of these characteristics are language-dependent.

Age estimation results suggest that listeners have detailed perceptual representations of vocal age. Certain speech characteristics are associated with talkers of a particular age or a particular age group (e.g. Ryan & Burk, 1974). Because people seem to know how the speech of young people should sound and how the speech of elderly speakers should sound, when speakers try to disguise their age, their age-disguised speech would presumably more or less match with the perceptual speech images that they have for a certain age groups (e.g. younger talkers or elderly talkers).
Age-related perceptual speech images could be realized if speakers project these images when they pretend to speak like the person in their image. In other words, when speakers are asked to disguise their age, they might manipulate the most representative speech characteristics of younger or older speakers based on the stored speech images they have for speakers of a certain age.

It seems that perceptual representation for vocal age is continuous in nature (Shipp & Hollien, 1969). Although the exact relationship between the perceptual representation of vocal age and the stereotype of certain age groups is not known, listener’s perceptual representations of vocal age should be affected by the stereotypical speech image for ‘younger’ or ‘older’ people during the listener’s age estimation process. Because stereotypical images of age are probably different in different cultures, stereotypical speech for age is probably different for Japanese and Americans.

In this chapter, perception of stereotypical age is explored. It is assumed that when speakers are asked to deliberately speak younger or older than their chronological age, their age-disguised speech would include some stereotypical characteristics of younger or older speakers. If so, then the question is how accurately listeners can perceive the age of talkers when they estimate the age of talkers who were instructed to disguise their age.

People can conceal their identity reasonably well when they speak in a disguised voice. However, it is known that listeners can identify individual speakers by listening to their speech when they know or have heard the speakers before. Accuracy of speaker identification is better when the listeners are more familiar with the speakers (Hollien, Majewski, & Doherty, 1982; Pollack et al., 1954; Van Lancker, Kreiman, & Emmorey,
Among various styles of speech disguise, it seems that age disguise is not very effective. Reich and Duke (1979) asked 40 male speakers (aged 21-42 years) to read sentences in five disguised voices and in a normal (undisguised) manner. In each disguised condition, the speakers were asked to disguise their utterances by changing their age (“70-80 year old speaker”), voice qualities (“severely hoarse voice” and “severely hypernasal voice”), speaking rate (“extremely slow rate”), and their identity. Each stimulus consisted of one undisguised utterance and one disguised utterance. Half of the stimulus pairs were spoken by the same speaker and another half of the pairs were spoken by different speakers. Listeners listened to each stimulus pair, and decided whether two utterances were spoken by the same or different speakers. Listeners could detect the disguised utterances in all five disguised conditions. When the listeners listened to the undisguised voice and the age disguised voice, they correctly detected the speaker identity with 67.6% accuracy (chance probability was 50%). When the listeners had extended knowledge about speech science, the accuracy for detecting disguised voices increased to 79.7%. Although the average success rate of disguise was not high, the listeners, even trained ones, failed to detect the identity of some of the talkers in the age-disguised condition.

Because Reich and Duke (1979) examined the effects of vocal disguises on the ability to identify speakers, it is difficult to apply their results to the perception of age for their age-disguised condition. First, the participants in their study were instructed to
imitate a “70-80 year old speaker” regardless of their actual age. Therefore, the difference between the disguised age and the actual age varied for individual talkers. Second, none of the talkers were elderly, and thus talkers in an age-disguised voice would not be confused with any other talkers. Hence, their results suggest that listeners sometimes decided that the age-disguised utterance was spoken by a different speaker, but it cannot be determined whether listeners perceived the age of age-disguised speech was as intended by the talkers, i.e. 70-80 years old.

It is also unknown whether elderly talkers are able to exhibit the same level of age-disguised performance as younger talkers. Age-related anatomical and physiological changes in the vocal tract could reduce the plasticity of vocal control among the elderly talkers. For example, muscular and morphological alterations in the larynx due to aging could impede elderly talkers from speaking in a very low or very high pitch. Irregular vocal fold vibration, which is one of the most common phenomena observed in elderly talkers (Ryan & Burk, 1974), may make it difficult for the elderly talker to make their voice quality smoother. Because the aging process is essentially irreversible, and because of the reduced plasticity of a vocal control, elderly talkers may not be able to change their voice to sound like younger talkers. Furthermore, because men are subject to earlier and more substantial changes in their voices than women, difficulties in manipulating their voice might be observed particularly in older male talkers.

In the present study, talkers were asked to read a sentence as if they were either twenty year older (Maturation condition) or twenty years younger (Rejuvenation condition) than their actual age. They also read the same sentence without disguise (Normal condition). The two age-disguised voices are termed the matured voice and
rejuvenated voice as compared to the normal (undisguised) voice. Listeners listened to a phrase “BCC” extracted from the sentences, and estimated the age of talkers. As a working assumption, rejuvenated or matured speech obtained from each talker will be considered as highly stylized speech that represents stereotypical vocal images of younger or older speakers. It was also assumed that an age difference of twenty between the chronological and disguised age is large enough to elicit stereotypical young speech or old speech so that each talker will exaggerate some speech characteristics in order to imitate a voice of younger or older speakers.

The previous forensic phonetics studies examined how accurately listeners can detect the speaker identity. Although perception accuracy was important in this study, unlike the previous forensic phonetics studies on speaker identification, the main focus here was not to find out how successfully listeners could identify the age of talkers in a disguised voice, but rather to examine whether the perceived age for age-disguised speech is the same or not despite of different indexical speech properties such as sex, age, and linguistic background.

**Purpose**

The main aim of the current study was to examine the effects of age disguise on age perception. Furthermore, this study examines whether highly stylized speech representing a vocal stereotype of old or young speakers is perceived in the same way by both native and non-native listeners. In addition to the effects of listener’s language on age perception of age-disguised speech, differences in talker’s age were examined to see whether age disguise by the younger talkers is better than age disguise by older talkers. Alternatively, older speakers, due to their level of experience, may have more detailed
age stereotypes. The effect of talker’s sex was also examined because men and women undergo different age-related changes over a different time course, and therefore presumably different stereotypical images are associated with men and women. Furthermore, we examined whether or not the direction of age-disguise would influence a listener’s perception of age-disguised speech because people may favor youth over seniors.

More specifically, the current study asked the following questions:

1. When listeners hear their native language, would the listeners increase or decrease the estimated age of the talker in age-disguised voice (matured or rejuvenated) relative to the same talker in a normal voice? Or more succinctly, can talkers successfully disguise their age or not?

2. When non-native listeners hear age-disguised speech and estimate the age of talkers, do they demonstrate the same age estimation accuracy as native listeners?

3. Is there any difference in age estimation accuracy between the two disguised conditions? For which conditions are listeners able to estimate the age of talkers more accurately? In other words, would the talkers be more successful in disguising themselves as their younger-self or their older-self?

4. Would females be more successful impostors than males, or vice versa?

5. Would Japanese speakers be more successful impostors than English speakers, or vice versa?

6. Would younger talkers be more successful impostors than older talkers?
Predictions:

A significant difference in age estimation between the rejuvenation condition and the normal condition is predicted. It was also expected that age estimation would be significantly higher for the maturation condition than the normal condition. However, the same results might not be obtained for both familiar and less-familiar languages because acoustic characteristics manipulated for age disguise in one language might not indicate stereotypical ‘youthful’ or ‘elderly’ of speech in other language.

Regardless of whether the listeners listened to a familiar language or a less familiar foreign language, perceived age differences between the normal and the age-disguised conditions would be smaller for the older group of talkers. This was expected for male talkers, especially for older male talkers, who might have a disadvantage in controlling their vocal structures to disguise their age due to physiological restrictions associated with aging. On the other hand, younger talkers might have a disadvantage in mimicking older speakers because they may have less experience with older talkers, and also have not experienced the older stage of life themselves. Then, age difference between the perceived age in the maturation condition and the perceived age in the normal condition might be smaller than the age differences of perceived ages between the normal and the rejuvenation conditions.

If older talkers do not have enough vocal plasticity to fully control their own voice due to age-related physiological changes, they might encounter some limitations in manipulating their speech in order to achieve a successful age-disguised voice. Then, the difference in perceived age between the age-disguised and normal conditions would be larger for the estimation for younger talkers than for older talkers.
The directionality of age-disguise was examined using the results for the middle-aged talkers because they could disguise their age in either direction without being extremely young or old. While there may be no difference when the middle-aged talkers speak as if they were younger or older, two other possibilities were considered. It might be easier for middle-aged talkers to project themselves as younger people than older people, if negative stereotypical images toward the elderly prevented talkers mimicking a role of an older person. Then, perceived age difference between the maturation and normal conditions would be smaller than the perceived age difference between the rejuvenation and normal conditions. On the other hand, identifying themselves as older people might be easier because the stereotypical images associated with the elderly are stronger and less variable than the ones attached to the younger people. In this case, the age differences between the normal and disguised conditions would be smaller for the rejuvenation condition than the maturation condition.

Methods

Talkers

Sixty talkers participated in the recording. Half of them were native speakers of American English who had been living in Bloomington, Indiana, USA. The other half were native speakers of Japanese, residents of Kobe, Japan. They were subdivided into three age groups: Young, middle-aged, and elderly groups. In each subgroup, there were an equal number of men and women. See Chapter 2 for the details of talkers.
Stimuli

The stimuli consisted of the phrase, a letter sequence “BCC”, that is pronounced in a similar way in Japanese and English. Each talker read the phrase embedded in a carrier sentence (see chapter 2) in their native language in the three conditions. First, they read the sentence in a normal manner. Then, they were asked to read the same sentence pretending that they were twenty-years older or twenty-years younger than they were. These age-disguised conditions are termed as maturation and rejuvenation conditions, respectively. Three utterances were recorded from each talker in each condition. The targeted phrase “BCC” was digitally edited as a stimulus from the second utterance of each talker in each condition. In order to avoid extreme unnaturalness, the utterances in the rejuvenation condition by young talkers (i.e. a 25 year old imitating a 5 year old) and the utterances in the maturation condition by the elderly talkers (i.e. a 85 year old imitating a 105 year old) were excluded from the stimulus set. There were 70 phrase stimuli (= 10 middle-aged talkers x 3 conditions + 20 young or elderly talkers x 2 conditions) for each language group of talkers; hence, in total, 140 phrase stimuli were used in the perception experiment.

Listeners

Twenty-four native speakers of American English and twenty-four native speakers of Japanese participated as listeners in the perception experiment. They were the same listeners participated in the experiment described in Chapters 3 and 4.
Procedure

The procedure was exactly same as described in Chapter 3. All the phrase stimuli were presented to each listener in a different random order, but they were blocked by talkers’ language. Listeners were told that all the stimuli were natural speech, but no other information about the stimuli was provided to the listeners in advance. Therefore, listeners did not know about the age-disguise, and they also did not know how many stimuli were produced by the same speaker.

Analysis

It was assumed that age-disguised speech would include stereotypical characteristics of younger or older speakers. If listeners successfully detected the speech differences introduced by age-disguise, then the perceived age for talkers in the age-disguised voice would be different from the age estimated for the same talker in the normal condition. In order to see whether the perceived age for each talker in the normal condition was different from that in the age manipulated condition (either rejuvenation or maturation), a value called age estimation difference (or Δage) was computed for each talker by subtracting the perceived age in the age-disguise condition (either the maturation or the rejuvenation) from the perceived age in the normal condition.

Responses for English talkers and Japanese talkers were analyzed separately. Within a language group, the talkers were subdivided into four groups by their age group and sex: middle-aged males, middle-aged females, elderly males, and elderly females in the analysis of rejuvenation condition; young males, young females, middle-aged males, and middle-aged females for analysis of the maturation condition.
Results

Separate analyses were conducted for each age-disguise condition. The results of rejuvenation condition will be presented first followed by the results of the maturation condition.

Rejuvenation condition

Table 28 summarizes the means and standard deviations of Δage for each talker-language group for each listener group. Both English and Japanese listeners exhibited similar mean values of Δage both for English talkers, and for Japanese talkers. Note that if the talker could successfully disguise their age as 20 years younger, then the expected Δage would be −20 years. Hence, although average perceived ages for most of the talkers were younger in the rejuvenation condition than the normal condition, the listeners did not perceive the rejuvenated speech as being as young as the talkers intended.

<table>
<thead>
<tr>
<th>Talkers</th>
<th>English listeners</th>
<th></th>
<th>Japanese listeners</th>
<th></th>
<th>All listeners</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>English</td>
<td>−2.54</td>
<td>15.84</td>
<td>−2.48</td>
<td>18.35</td>
<td>−2.51</td>
<td>17.13</td>
</tr>
<tr>
<td>Japanese</td>
<td>−5.18</td>
<td>19.88</td>
<td>−4.90</td>
<td>18.01</td>
<td>−5.04</td>
<td>18.96</td>
</tr>
<tr>
<td>Average</td>
<td>−3.86</td>
<td>18.02</td>
<td>−3.69</td>
<td>18.21</td>
<td>−3.77</td>
<td>18.11</td>
</tr>
</tbody>
</table>

Effect of age disguise and listener language

Speech in youthfully disguised voice (rejuvenated speech) was overall perceived as younger than the undisguised speech. However, the values of Δage were considerably smaller than the expected value of Δage (i.e. −20). In order to see whether the perceived
Table 29. Two-way ANOVAs on perceived age estimated by English and Japanese listeners (L) for talkers in the rejuvenation and normal conditions (C). Four separate ANOVAs are shown for the English middle-aged and elderly talker groups, and Japanese middle-aged and elderly talker groups.

<table>
<thead>
<tr>
<th></th>
<th>English talkers</th>
<th>Japanese talkers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Middle-aged</td>
<td>Elderly</td>
</tr>
<tr>
<td></td>
<td>Middle-aged</td>
<td>Elderly</td>
</tr>
<tr>
<td>Listener Language (L)</td>
<td>0.40</td>
<td>2.77</td>
</tr>
<tr>
<td></td>
<td>3.20</td>
<td>23.55*</td>
</tr>
<tr>
<td>Error</td>
<td>478 (228.87)</td>
<td>(380.34)</td>
</tr>
<tr>
<td></td>
<td>(466.98)</td>
<td>(403.10)</td>
</tr>
<tr>
<td>Condition (C)</td>
<td>1.56</td>
<td>26.23*</td>
</tr>
<tr>
<td></td>
<td>34.01*</td>
<td>33.69*</td>
</tr>
<tr>
<td>C × L</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.54</td>
<td>0.21</td>
</tr>
<tr>
<td>Error</td>
<td>478 (141.20)</td>
<td>(150.84)</td>
</tr>
<tr>
<td></td>
<td>(192.34)</td>
<td>(168.01)</td>
</tr>
</tbody>
</table>

Note. Values in parentheses represent mean square errors. *p < .001.

The mixed two-way ANOVA results for each of the four talker groups (English middle-aged, English elderly, Japanese middle-aged, and Japanese elderly talkers) are summarized in each column in Table 29. The main effects of Age-disguise were significant for English elderly talkers and Japanese elderly talkers. As for the middle-aged talkers, only Japanese talkers were perceived younger for their rejuvenated speech relative to the normal speech, while English middle-aged talkers showed no significant effect of age-disguise. The effect of listener’s language was not significant for most of
the talker groups except the Japanese elderly talker group. Japanese listeners perceived the rejuvenated speech by the Japanese elderly talkers as older than English listeners did.

Table 30 lists the average perceived ages for each of the four talker groups (English middle-aged, English elderly, Japanese middle-aged, and Japanese elderly talkers) in the rejuvenation and normal conditions. The average perceived ages by the English listeners are found in the middle two columns, and the average perceived ages by the Japanese listeners are in the two right-most columns. It seems that the difference between the two listener groups for the Japanese elderly talkers was due to the fact that Japanese listeners judged the elderly Japanese as older than the English listeners did in both normal and rejuvenation conditions, as can be seen in the bottom row in Table 30. However, the differences in the perceived ages in two conditions were very similar in both listener groups as is evident in the lack of interaction between listener and condition [Δage for English and Japanese listeners = 4.5 and 5.5 years, respectively].

Table 30. Average perceived age in the rejuvenation and normal conditions in years.

| Talkers | Conditions     | English listeners | | | | Japanese listeners | | | |
|---------|----------------|------------------|---|---|---|------------------|---|---|
|         |                | Perceived Age    | Rejuvenation | Normal | Rejuvenation | Normal | |
| English | Young          | NA               | NA           | NA     | NA           | NA     | |
|         | Middle-aged    | 40.0             | 40.7         | 39.2   | 40.3         | |
|         | Elderly        | 52.2             | 56.5         | 54.5   | 58.3         | |
| Japanese| Young          | NA               | NA           | NA     | NA           | NA     | |
|         | Middle-aged    | 45.7             | 51.6         | 43.9   | 48.4         | |
|         | Elderly        | 61.1             | 65.6         | 67.0   | 72.3         | |

Note: Average perceived age for young talkers in the normal condition are not presented because there are no comparable values in the rejuvenation condition.
Effects of talker sex and talker age

In order to see the effects of talker’s sex and age, the results of ∆age were divided further into eight subgroups (i.e. 2 talker languages × 2 talker sex × 2 age groups).

Three-way ANOVA tests were performed on the ∆age data with two levels of listener’s language, two levels of talker’s sex and two levels of talker’s age group as fixed factors.

The three-way ANOVA was conducted separately for the English talker group and the Japanese talker group. The summary of the ANOVA results is shown in Table 31.

Figure 13 displays the mean ∆ages for the English talkers in the English listener group (solid) and the Japanese listener group (striped). Figure 14 shows the mean ∆ages for the Japanese talkers. In both figures, talkers were grouped by their age group and sex.

Table 31. Results of two separate three-way ANOVAs for listener’s language (L), talker’s sex (S), talker’s age group (A) on ∆age (= Perceived age in rejuvenation condition − Perceived age in normal condition) for each talker.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>English talkers</th>
<th>Japanese talkers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listener Language (L)</td>
<td>1</td>
<td>&lt; 0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Talker Sex (S)</td>
<td>1</td>
<td>4.34</td>
<td>4.42</td>
</tr>
<tr>
<td>Talker Age (A)</td>
<td>1</td>
<td>8.08*</td>
<td>0.09</td>
</tr>
<tr>
<td>L × S</td>
<td>1</td>
<td>.34</td>
<td>0.53</td>
</tr>
<tr>
<td>L × A</td>
<td>1</td>
<td>.12</td>
<td>0.72</td>
</tr>
<tr>
<td>S × A</td>
<td>1</td>
<td>20.01**</td>
<td>0.51</td>
</tr>
<tr>
<td>L × S × A</td>
<td>1</td>
<td>.29</td>
<td>1.57</td>
</tr>
<tr>
<td>Error</td>
<td>952</td>
<td>(285.77)</td>
<td>(359.21)</td>
</tr>
</tbody>
</table>

Note. Values in parentheses represent mean square errors.  
* p < .01.  ** p < .001.
Figure 13. Differences in the perceived age for English talkers between the rejuvenation and normal conditions.
The three-way ANOVA results for the English talkers showed a significant effect of age group, but there was no significant difference in terms of talker’s sex. There was only one significant interaction effect for the talker’s sex and talker’s age group. A Tukey post-hoc test revealed that the elderly English female talkers were significantly different from other groups of English talkers as seen in Figure 13. Regardless of the listener’s language, elderly American women disguised their age more successfully among the English talker groups by reducing their perceived age about 7 years. Indeed, the significantly large $\Delta$age found for the elderly American women was consistent with

Figure 14. Differences in the perceived age for Japanese talkers between the rejuvenation and normal conditions.
the analysis on the perceived age above (see Table 29), where the perceived age for
English elderly talkers exhibited a significant difference between the two conditions.

Turning to the results for the Japanese talkers, the three-way ANOVA results
showed no significant main or interaction effects. However, unlike the English talkers,
Δage for the Japanese talkers were consistent (about ~5 years) regardless of talker’s sex
and age and across the listener groups. When we compare Figure 13 and Figure 14, it
seems that the mean Δage values for English talkers were smaller than the mean Δage for
Japanese talkers except English elderly women. Hence, Japanese talkers overall reduced
their perceived age in their rejuvenation condition successfully relative to English talkers.

To summarize, perceived age in the rejuvenation condition was younger than the
perceived age in the normal condition although the effect of age-disguise was small in
terms of Δage values. Listener’s native language, talker’s age, and sex did not influence
the Δage in general except the perception of rejuvenated speech by the English elderly
women.

**Maturation condition**

The means and standard deviations of Δage for each talker group are displayed in
Table 32. When a listener’s perceived age for a given talker in a matured voice was
older than the age perceived for the normal voice, Δage will be a positive value. Positive
values of mean Δage were observed for both English and Japanese talkers in both listener
groups (see Table 32). However, the differences in perceived age between the two
conditions were small (English talkers: $M = 5.5$ years; Japanese talkers: $M = 2.5$ years)
relative to the intended age difference of 20 years, similar to the small difference in
perceived age for the rejuvenation condition. This may be partially explained by
variability in Δage among the talkers. Mean Δages for individual talker (among all listeners) ranged −6.6 to 34.0 years for English talkers, and −15.8 to 18.4 years for Japanese talkers. Unexpectedly one Japanese female talker in the middle-aged group (JFM01) exhibited a reverse effect on perceived age for her disguised speech of −15.8 years. If her results are excluded, the lower end of mean Δage for Japanese talkers is −4.8, which is similar to the lowest mean Δage for English talkers. Even so, it is noteworthy that some of the speakers were perceived as younger by the listeners when they tried to speak as if being 20 years older.

Table 32. Means and Standard Deviations of Δage calculated as maturation age minus normal age in years.

<table>
<thead>
<tr>
<th>Talkers</th>
<th>English listeners</th>
<th>Japanese listeners</th>
<th>All listeners</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>English</td>
<td>6.38</td>
<td>16.00</td>
<td>4.65</td>
</tr>
<tr>
<td>Japanese</td>
<td>2.63</td>
<td>18.94</td>
<td>2.35</td>
</tr>
<tr>
<td>Average</td>
<td>4.51</td>
<td>17.62</td>
<td>3.50</td>
</tr>
</tbody>
</table>

Effect of age disguise

In order to see whether the perceived age differences between the maturation condition and the normal condition are significant, four mixed two-way ANOVAs were conducted on perceived age, with the age disguise condition (C) as the within variable and the listener language (L) as the between variable. The effect of listener language was tested to see whether both English and Japanese listeners perceived the disguised age in the same way or not. Separate analyses were conducted for each age group (Young and Middle-aged) and each language group of talkers (English and Japanese talkers). Results
of the four two-way ANOVAs (Age-disguise condition × Listener language) are presented in Table 33.

Considering the age-disguise condition (C), results of the ANOVA revealed that speech in a matured voice was perceived as older than undisguised speech for three of the talker groups, middle-aged English, middle-aged Japanese, and young Japanese talkers. However, perceived age was not significantly different between maturation and normal condition for young English talkers.

### Table 33. Two-way ANOVAs on perceived age estimated by English and Japanese listeners (L) for talkers in the rejuvenation and normal conditions (C). Four separate ANOVAs are shown for English young and middle-aged talker groups, and Japanese young and middle-aged talker groups.

<table>
<thead>
<tr>
<th></th>
<th>English talkers</th>
<th>Japanese talkers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>df</td>
</tr>
<tr>
<td>Between variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listener Language (L)</td>
<td>1</td>
<td>0.03</td>
</tr>
<tr>
<td>Error</td>
<td>478</td>
<td>(198.21)</td>
</tr>
<tr>
<td>Within variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition (C)</td>
<td>1</td>
<td>2.78</td>
</tr>
<tr>
<td>C × L</td>
<td>1</td>
<td>1.99</td>
</tr>
<tr>
<td>Error</td>
<td>478</td>
<td>(82.80)</td>
</tr>
</tbody>
</table>

**Note.** Values in parentheses represent mean square errors. *p < .05. **p < .01. ***p < .001.

### Effect of listener language

The effect of listener language (L) was not significant for English talkers, but was significant for Japanese young talkers, and marginally significant for Japanese middle-aged talkers. None of the interactions were significant. Results for the middle-aged talkers are of particular interest, because for both English and Japanese talkers
significantly higher perceived age was found for age-disguised matured speech with respect to the perceived age for undisguised normal speech. For comparison, the mean perceived ages in the two conditions are listed side-by-side for each talker group in Table 34. English middle-aged talkers (highlighted in grey) were perceived older in the maturation condition than the normal condition by both listener groups, on average 10 years. On the other hand, perceived age for the Japanese middle-aged talkers in the age-disguised condition was only about three years older than their perceived age in the normal condition. In both conditions, Japanese middle-aged talkers were perceived older by English listeners than by Japanese listeners. Hence, for both English and Japanese listeners, English middle-aged talkers disguised their age more successfully than Japanese middle-aged talkers. Moreover, the results suggested that the listener language did not have a great effect on perceived age of disguised speech.

<table>
<thead>
<tr>
<th>Talkers</th>
<th>Conditions</th>
<th>English listeners</th>
<th>Japanese listeners</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Maturation</td>
<td>Normal</td>
</tr>
<tr>
<td>English</td>
<td>Young</td>
<td>32.2</td>
<td>30.4</td>
</tr>
<tr>
<td></td>
<td>Middle-aged</td>
<td>51.7</td>
<td>40.7</td>
</tr>
<tr>
<td></td>
<td>Elderly</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Japanese</td>
<td>Young</td>
<td>41.2</td>
<td>38.9</td>
</tr>
<tr>
<td></td>
<td>Middle-aged</td>
<td>54.5</td>
<td>51.6</td>
</tr>
<tr>
<td></td>
<td>Elderly</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: Average perceived age for elderly talkers in the normal condition are not presented because there are no comparable values for the elderly talkers in the maturation condition.
**Effect of talker sex and talker age**

In order to examine whether there is any difference in the ability to disguise voice between female and male talkers, two separate three-way ANOVAs were conducted on ∆age of each talker for each listener group with two levels of listener language (L), two levels of age group (A) (young and middle-aged talkers), and two levels of talker’s sex (S) as fixed factors. **Figure 15** and **Figure 16** display the mean ∆ages in English listeners (solid bars) and Japanese listeners (striped bars) for English and Japanese talkers, respectively. In both figures, talkers were sub-grouped in terms of their sex and age group. A summary of the two three-way ANOVA results is found in **Table 35**.
Figure 15. Differences in the perceived age for English talkers between the maturation and normal conditions.
Figure 16. Differences in the perceived age for Japanese talkers between the maturation and normal conditions.

Table 35. Results of two separate three-way ANOVAs for listener’s language (L), talker’s sex (S), and talker’s age group (A) on Δage (= Perceived age in maturation condition – Perceived age in normal condition) for each talker.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>English talkers</th>
<th>Japanese talkers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listener Language (L)</td>
<td>1</td>
<td>2.63</td>
<td>0.06</td>
</tr>
<tr>
<td>Talker Sex (S)</td>
<td>1</td>
<td>13.02*</td>
<td>1.93</td>
</tr>
<tr>
<td>Talker Age (A)</td>
<td></td>
<td>72.35*</td>
<td>1.45</td>
</tr>
<tr>
<td>L × S</td>
<td>1</td>
<td>0.42</td>
<td>0.01</td>
</tr>
<tr>
<td>L × A</td>
<td>1</td>
<td>0.00</td>
<td>0.35</td>
</tr>
<tr>
<td>S × A</td>
<td>1</td>
<td>0.01</td>
<td>1.74</td>
</tr>
<tr>
<td>L × S × A</td>
<td>1</td>
<td>0.31</td>
<td>4.59</td>
</tr>
<tr>
<td>Error</td>
<td>952</td>
<td>(273.21)</td>
<td>(259.64)</td>
</tr>
</tbody>
</table>

Note. Values in parentheses represent mean square errors.
* p < .001.
Examining Δage for English talkers in Figure 15, young talkers were clearly poorer at matured age disguise than middle-aged talkers. The main effect of age group was significant, but the listener language effect was not significant. This means that both English and Japanese listeners found that middle-aged talkers were better at disguising their age than young talkers. The effect of talker sex was also significant. Male English talkers disguised age better than female English talkers. Other comparisons for English talkers, including all interactions, were not significant. In the analysis of Japanese talkers, none of the main and interaction effects were significant. Despite that the perception results overall suggested that the Japanese listeners estimated age of talkers more accurately than the English talkers, it seems that the results of age perception for the maturation speech indicated that Japanese talkers were very poor at disguising themselves as older persons.

Discussion

Talkers

When the talkers spoke as if they were twenty years younger than their actual age, their disguised speech sounded younger to the listeners. However, the middle-aged talkers of English were not perceived as younger in their rejuvenated speech compared to the perceived age for their normal speech. When the talkers disguised their age as older than themselves, the listeners’ perceived age for the matured speech was older than the perceived age for the normal speech, except that the perceived age for the young talkers of English did not change between the two conditions. The reason for these two exceptional cases in the results for English talkers is not clear, but may be related to the tendency for age overestimation for young talkers and for age underestimation for
middle-aged talkers in perceived age in the normal condition. As a consequence of these tendencies for an offset of age estimation for the normal speech, the listeners would need to estimate the age of thirties through early fifties with an extraordinary accuracy to adjust for the offsets. Regarding the English talkers, there was no difference between English and Japanese listeners in perceiving age-disguised speech. Hence, it might be that the age of rejuvenated speech by middle-aged English talkers and the age of matured speech by young English talkers were simply more difficult to estimate, i.e., disguised age did not incorporate cues that signaled an actual age.

Although overall results showed that age-disguised speech was perceived older in the maturation condition and younger in the rejuvenation condition than in the normal condition, the effects of age-disguise were small. This may be due to unnaturalness that the listeners might have perceived in the age-disguised speech. If the age-disguised speech sounded ‘fake’ to the listeners, the listeners may adjust their estimated age accordingly.

**Listener’s language**

Results revealed that the listener’s language has limited influence on the age estimation of age-disguised speech. Japanese listeners did estimate the age of talkers more accurately in the normal condition than English listeners in general. However, the perceived age difference between the normal and either of the age-disguised conditions was essentially the same in both listener groups. This suggests that both English and Japanese listeners might have similar stereotypes of speech associated with a certain age group. Results also imply that Japanese listeners did not have an advantage in age estimation for the age-disguised speech. Remember that the Japanese listeners were
better judges of age estimation than the English listeners. If the Japanese listeners estimated the age of talkers in the age-disguise conditions using the same perceptual cues that they employed in age estimation in the normal condition, then, the $\Delta$age exhibited by the Japanese listeners could be larger than the $\Delta$age in the English listeners if the talker’s disguise were successful. Alternatively, the $\Delta$age in the Japanese listeners could be smaller if the listeners were not fooled by the talker’s disguise. Apparently, age-disguise did not create a great impact on listener’s age judgments.

**Age of talkers**

We predicted that older talkers (the elderly and middle-aged groups) might not be as successful as the younger talkers when they changed their voice to be younger or older than themselves. In particular, we predicted that elderly male talkers might have more difficulty in disguising their vocal age than the female elderly because men tend to show earlier and more substantial physiological changes due to advancing aging than women. However, the outcome of this research does not support these predictions. The elderly talkers showed similar or larger values of $\Delta$age than the middle-aged talkers in the rejuvenated speech. Likewise, in the matured speech, the middle-aged talkers showed similar or larger $\Delta$ages than the young talker groups. Some sex differences were observed. The elderly English women exhibited larger age differences between the rejuvenation and normal conditions than the elderly English men. Japanese female talkers in the elderly group seemed to show a slightly larger size of $\Delta$age for rejuvenated speech than Japanese elderly men, but the difference was not significant. In the maturation condition, there was no $\Delta$age difference in terms of talker sex in the middle-aged talkers for English or Japanese. Therefore, when both male and female talkers
change their voice to disguise their age, listener’s perceptual judgments of age are not influenced by the talker’s sex. Although some of the young talkers said to the investigator that their speech would not change in twenty years, the experimenter noticed that the young talkers did try to change their speaking style in the maturation condition pretending as if they were 45-50 years of age. However, the age-disguise by young English talkers was obviously not good enough to convince the listeners.

There are at least two ways to account for these age-related results depending on who receives credit for a small size effect of age-disguise on age perception, the talkers or listeners. The extreme listener-oriented view considers that the listeners outperformed the talkers. Although the talkers tried to disguise their age, they were not able to conceal the properties of young speech so that the cleverer listeners still picked up the real age identity of talkers, even in the disguised speech. Alternatively, the extreme talker-oriented view would be explained by the poor quality of talker’s disguise due to talker’s unsuccessful strategies. That is, the talkers manipulated some speech/voice features, but these features did not match the speech characteristics associated with middle-aged talkers. The later view implies that the stereotypical speech characteristics that young talkers attached to the middle-aged talkers do not match the perceptual speech characteristics they relied on for age estimation of middle-aged talkers. Which of the two views is right cannot be determined by the current study. It might be possible to test these extreme hypotheses by employing professional actors and forensic experts who specialized on speaker identification. However, more realistically, the current results would be explained by a detailed acoustic and perceptual analysis for individual talkers.
Age of listeners

The age of the listeners could influence their age perception (Huntley et al., 1987). The listeners in the current study were all young (mean age = 19.8 years). They were in fact fairly accurate in estimating the age of middle-aged talkers in the normal condition, which suggests that young people have a concept of how people in forties should sound like. The young talkers in this study were slightly older than the listeners, but their ages were similar (mean age = 26.7 years). However, given that the young talkers in this study also have the same perceptual image for the talkers in forties as the young listeners, the young talkers appeared unable to match those perceptual stereotypes to their own speech production in the disguise condition. Young people who face and interact with a lot of people older than their age cohorts might be able to realize the fact that aging is an inevitable process for everyone including themselves. The young listeners in this study might not have had extensive interactions with the middle-aged and elderly people. The results might be different if the young listeners were selected because they interacted with a lot of middle-aged and elderly people. Furthermore, whatever stereotypes one has for young or old people, those stereotypes could change as one becomes older. Hence, it would be interesting to examine how middle-aged listeners would perceive the age of speech with and without age-disguise.

Acoustic properties for older voices

The results of this study also lead to the interesting question of what kinds of information the talkers manipulate when they tried to change their age. Some acoustic properties must be either enhanced or reduced in order to match their speech to fit to whatever auditory stereotypes each talker had for younger or older speakers. Preliminary
acoustic analysis on the age-disguised speech by the middle-aged talkers indicates that the middle-aged talkers increased their pitch and speaking rate when they spoke in the rejuvenated condition (Nagao, 2005). These results support the findings in the previous literature that perceived age is correlated with average fundamental frequency (Hartmann & Danhauer, 1976; Horii & Ryan, 1981; but see Braun & Rietveld, 1995) and speaking rates (Braun & Rietveld, 1995; Ptacek & Sander, 1966; Shipp et al., 1992; Ryan & Burk, 1974). However, when the matured speech by the middle-aged talkers was compared with their normal speech, there was no significant change in average fundamental frequencies and average duration in their matured speech.

The acoustic results of age-disguised speech suggest that the talkers failed to project the actual age-related speech characteristics. In particular, despite the fact that a slow rate of speech is listed as one of the most obvious characteristics of the speech by the elderly, it is surprising that half of the middle-aged talkers did not slow down their speech when they tried to speak as if they are 20 years older than their actual age. It is known that when the talkers imitate someone’s voice, they tend to assimilate their average pitch level to the average fundamental frequency of the person they impersonate (Endres, Bambach, & Flosser, 1971). Endres et al. (1971) also reported that fundamental frequencies that an imitator manipulates do not exactly match with the actual fundamental frequencies of the imitated person. Almost no changes in average fundamental frequencies between normal and matured speech by the middle-aged talkers suggest that average pitch may not be closely related with stereotypical speech associated with the elderly talkers among the middle-aged talkers. Nevertheless, the matured age-disguise had an intended effect on their perceived age. It is possible that other acoustic
properties were manipulated in matured voice. It remains to be answered whether the stereotypical speech image for particular age groups is merely imaginary, or based on the same acoustic properties employed for estimating age of talkers with some properties either enhanced or reduced its perceptual contributions.

Preliminary acoustic analysis indicated that the age-disguised speech did not exhibit some of the obvious acoustic changes in speech which are typically found due to physiological aging. It seems that the listeners have at least two representations associated vocal age: one for speech representations with detailed acoustic information, which is more closely associated with chronological age, and another for the speech representation for age stereotypes, which is more socio-culturally bound and could reflect acoustic characteristics evoked by the social images and attitudes toward the elderly and the young people in each linguistic community.

Conclusions

In this chapter, the effect of vocal disguise on age perception was examined. Results demonstrated that the age of talkers was perceived younger or older than their actual age in line with the talkers’ intended direction of age shift in their age-disguised voice. However, the effect of age-disguise was generally small so that their disguised speech was still heard as people within the boundaries of their own age groups. Age-disguise influenced listeners’ age estimation similarly for both English and Japanese listeners. Current findings need to be evaluated carefully with acoustical analysis because different acoustic factors can provide the same perceptual effects. Nonetheless, it seems that in both Japan and the United States, the stereotypical speech images for younger or older talkers are similar among the young listeners in this investigation.
6. General Discussion and Conclusion

Summary of Results

The current study examined cross-language effects in the perception of a talker’s age. Relationships between the listeners’ perceived age and language familiarity were examined using three kinds of stimuli that included different amounts of linguistic information. Correlations between chronological age and perceived age for talkers became higher as the amount of information increased (vowel < phrase < sentence). In terms of the effect of language familiarity, the results were different depending on the type of stimuli. When listeners listened to the vowel stimuli, moderate correlations were found between the talker’s chronological age and the listener’s perceived age regardless of whether the listeners judged the stimuli spoken in the familiar language or less familiar language. However, a listener’s own language compared to a less-familiar language clearly showed an influence on the age estimation when more linguistic information was available. That is, when the phrase and the sentence stimuli were presented to the listeners, the listeners showed better age estimation for the talkers in the familiar language than the talkers of the foreign language.

In addition to the correlation results, age estimation accuracy was examined. Age estimation errors decreased as the amount of linguistic information in the stimuli
increased. Age estimation was quite accurate for the young talkers. However, the listeners showed a tendency to underestimate the age for middle-aged and elderly talkers. The mean perceived age was 13 years younger than the mean chronological age for the middle-aged talkers and 20 years younger for the elderly talkers. Furthermore, when age estimation was made for the talkers in a familiar language, Japanese listeners were more accurate judges than English listeners.

Additional analyses examined the effect of talker’s sex on the listener’s perceived age for the sentence stimuli. Correlations were stronger between perceived age and chronological age for female talkers than male talkers. However, the effect of talker’s sex was only observed in the estimation for talkers in a familiar language. These results were best explained by the sociolinguistic differences between men and women. Although the investigation was limited to the data collected from the English listeners, the effect of listener’s sex on age estimation was also examined. There was no difference between English male and female listeners in terms of correlation between chronological age and perceived age in each context. Hence, a peer effect related to sex was not found in this study.

The current study also explored the perception of age-disguised speech in an attempt to see whether there is a cross-language difference in stereotypical speech for a certain age group. A language effect between the English and Japanese listeners was not found in the age perception of the disguised speech. Whatever strategies talkers adapted to disguise their age, the effect was the same for both English and Japanese listeners in terms of the perceived age differences between the disguised and normal conditions.
Rather unexpectedly, the perceptual effect of age-disguise was found to be quite small and not affected by language familiarity.

**Perceptual mechanisms in identification of talker’s age**

Given the results that listeners can estimate the age of talkers whether the talker speak listener’s native language or a foreign language, the question of what is the basis for the listeners’ judgments is raised. Speech sounds can convey various kinds of information. Information about a talker’s age is just a small part of talker-specific information. Talker-specific information means idiosyncratic characteristics that a listener can use to identify a specific individual, but also includes sociolinguistic (or indexical) information which enables a listener to identify the general linguistic backgrounds of the talker such as sex, dialect, and social class. It has been shown that listeners store detailed information about individual speakers (Mullennix, Pisoni, & Martin, 1989; Palmeri, Goldinger, & Pisoni, 1993; Pisoni, 1997; Mullennix, Johnson, Topcu-Durgun, & Farnsworth, 1995; Van Lancker et al., 1985a, 1985b) as well as their linguistic backgrounds such as sex (Lass, Hughes, Bowyer, Waters, & Bourne, 1976; Bennett & Montero-Diaz, 1982), ethnicity (Thomas & Reaser, 2004), and dialects (Clopper, 2004). Yet, the exact nature of perceptual representation of age is not clear. The process of perceptual estimation for talker’s age is discussed in this section. Four models are proposed and evaluated here.

**Exemplar-based model**

The central assumption of an exemplar model is that a listener stores individual exemplars as perceptual representations in long-term memory, and categorizes an item by
comparing its perceptual similarity with the stored exemplars (e.g. Nosofsky, 1992; Goldinger, 1996). An exemplar-based model for age perception assumes that people store age information as one of the attributes of speaker-specific characteristics.

Perception of a talker’s age could require direct access to the listener’s stored information for specific speakers. There are two steps in the age estimation process. When a listener estimates the age of an unfamiliar talker, it is assumed that the listener first searches for stored representations of familiar talkers who sound similar to the unfamiliar talker’s voice (Matching process). Searching through the stored speech representations of the listener’s familiar talkers to find the best match to the unfamiliar talker, the listener then needs to extract the age property from individual speech representations that best matched with the unfamiliar talker’s speech. A numeric value is given to the person based on the age associated with the familiar talkers (Association process).

**Prototype-based model**

In prototype models, it is assumed that people store the prototypes of categories. Prototypes are abstraced representations of certain categories, which contain average or ideal characteristics of all category members (e.g. Reed, 1972; Kuhl & Iverson, 1995; see for review, E. E. Smith & Medin, 1981; Medin & Barsalou, 1987). Age estimation in a prototype model is similar to the exemplar model, and involves two steps: categorization and age association. In a prototype-based model, age information can be encoded as a separate attribute of talker information. A listener has prototypical sounds\(^5\) for a certain age or age group based on the speech characteristics common to speakers in the same age

---

\(^5\) Prototypes can be similar to stereotypes, but they are not exactly the same. Prototypes are based on speech samples that the listener actually experienced, while stereotypes can be constructed without empirical instances.
or age group. When the listener hears an unknown talker, the listener selects the perceptually most similar prototype to the unknown talker (categorization process). Then, the listener assigns a numeric value based on the perceptual distance from the selected prototype as well as the perceptual distance from the neighboring prototype (association process). **Figure 17** is a schematic illustration of prototypes in perceptual space. In the categorization process, a new talker (represented as a smiley face) is classified as a member of prototype C because of the perceptual proximity to the prototype C over other prototypes. Perceived age for the talker (Mr./Ms. Smiley) is determined by computing the difference of distances from the age assigned to each prototype. When the assigned ages for prototype B and C in **Figure 17** are 40 and 60 years old, the age for the new talker is computed as 55. Although actual chronological age of Mr./Ms. Smiley is 72 in this example, the age estimation error occurred because the ages assigned to the prototypes B and C were not accurate (e.g. 50 and 80 years old).
Although we do not know what kinds of category prototypes are represented, prototypes are assumed to be rank-ordered in a multidimensional metric space. To differentiate the perceived age as a listener’s response and the perceptual representations of age, we will call the perceptual representations for ‘age’ as vocal age. Listeners’ age underestimation for older talkers can be explained by an error in the association process. An age association error could decrease accuracy, but not decrease correlation. When vocal age is accurately categorized, a high correlation can be obtained between chronological age and perceived age.

Figure 17. Age perception in a prototype-based model.
One concern about the prototype based model is that we do not know what kinds of prototypes listeners use when the task is direct age estimation. If the task is simple age classification, the categories for prototypes (e.g. young and elderly) that the listeners have could actually match with the category responses used in the task. That is, when the listeners identify the age of a talker as a ‘young talker’ or an ‘elderly talker’, the listeners might have a prototype corresponding to each of the ‘young talker’ and ‘elderly talker’ categories. However, there are a number of categories the listener could employ such as middle-age, early middle-age, very young, very old, twenties, thirties, mid-twenties, early twenties, eighty-something, etc. Although it may seem empirically unlikely, conceptually listeners could have a prototypical example for each age, e.g. a prototype of 80-year-old that is different from a 79-year-old. This might be true especially in direct age estimation because listeners identify a person’s age, rather than classify a person into some age groups. Presumably, there are some superordinate categories in such a case, e.g. prototypes for ’70-year-old’ and ’71-year-old’ are nested within ‘old people’ prototypes. Obviously, there are many questions unanswered about how the concept of age might be structured.

The nature of age category is also related with another question: how can the prototype-based model handle the difference between female and male speech. Prototypes for male and female speakers can be different or gender may simply be a different set of prototypes for an age. The male-female coalescence theory predicts the number of prototypes for elderly people is somehow reduced. That is, speech characteristics of elderly men and women become similar to each other with advanced aging (Hollien, 1995). If prototypes are not gender-specific, a listener would assign the
same perceived age for the best exemplar of elderly male talkers and the best exemplar of elderly female talkers. If prototypes are gender-specific, the number of prototypes would be almost doubled.

The same issue can be argued with regard to the effect of language familiarity on perceived age. The present results raise the question of how to model age estimation for foreign talkers. There are two possibilities in this model. When the same prototypes are used to identify the age of talkers whether the talker speaks in the native language or a foreign language, the prototype model predicts that listener’s age estimation is better for native speakers than foreign talkers. This is because the vocal age for native speakers is more similar to the listener’s prototypes than the vocal age for foreign speakers because prototypes are almost exclusively based on experience in the native language. Hence, error in the classification process accounts for the age estimation difference between the native speakers and non-native speakers. Alternatively, the listener could have some age prototypes (or exemplars) for foreign talkers besides the prototypes (or exemplars) for the native language. It is expected that the more prototypes listeners have, the more accurate perceived age will be. Presumably, listeners have a smaller number of prototypes for foreign speakers than native language; hence, the model predicts reduced age estimation for foreign speech. In other words, errors are associated with an insufficient number of prototypes.

As the reader may have noticed, as the number of prototypes increases, the prototype-based model becomes more similar to the exemplar model. Further investigation will be needed to answer the questions of how many and what kinds of
prototypes the listener stores and accesses to retrieve the age information and what kinds of cues form prototypes.

**Anchor model**

The current study proposes a new model, the anchor model, which is similar to the exemplar model in some ways, but similar to the prototype model in other ways. In this model, the age of an unknown talker is determined by relative comparison with persons who are very familiar to the listener. The listener uses someone as a reference, someone that the listener knows very well including their actual age. Since the listener knows the age of these referential persons, these persons will be used as buoys to anchor listener’s age perception. Using a relatively small number of people as references, the listener approximates the age distance between the unknown talker and these familiar persons. Since the listener knows the actual ages of these referential persons, the age of the talker is determined relative to the age of these people. Some of the best referential figures are family members of the listener.

Consider an example where a listener uses two family members as reference in order to estimate the age of an unknown talker. First, a listener selects one familiar person, e.g. his father at age of 55. Then, the listener decides whether this unknown talker is younger or older than his father. If the listener estimated the talker as younger than his father, then he compares the talker with another familiar person who is younger than his father, say his older brother whose age is 35. Now the listener guesses how much younger or older the talker is compared to his father and his brother, such as “he sounds older than my brother, but doesn’t sound like my father” or “he is definitely
younger than my father”. Hence, the listener estimates the age of the talker at somewhere between 35 (his brother’s age) and 55 (his father’s age) and relatively closer to his brother’s age. So, the listener estimates the talker’s age as 40. Accuracy improves if the listener employs more people for a comparison, while accuracy decreases if the listener uses fewer people for reference. Age estimation becomes more difficult and error prone if the age of referential persons is distributed in a narrow range.

This model is similar to the exemplar model in a sense that all talker characteristics can be encoded holistically. A difference from the exemplar and prototype models is that the listener’s judgment can rely on the persons for whom the listener is certain about their actual age. However, unlike the exemplar model, the listener’s age perception is not based on all speakers that the listener knows, just the most familiar persons. Moreover, the referential persons that a listener relied on for their judgments are not necessarily the good matches for persons at a certain age. One problem in the prototype model is that how to determine typicality for age prototypes. However, this model does not have this problem for the issue of typicality because the best models (prototypes or exemplars) are defined by the listener’s familiarity.

**Abstractionist model**

An abstractionist type approach (e.g. Tenpenny, 1995), is different from the three models discussed above. In this model, age information is treated as an independent category dissociated with any other speaker information. Age information is a completely independent perceptual category at some level of the recognition process.
In an abstractionist type model, a listener applies a single rule to compute the speaker’s age. Similar to traditional linguistic theory\(^6\), the age estimation rule can reduce the amount of information that one needs to store because age information will be available by applying the rule. The age estimation rule is context-free and language universal. The age estimation rule can be based on the physiological model of vocal aging like the theory that Hollien (1995) proposed. Hollien (1995) proposed the physiological based model of vocal age, where vocal age is associated with several acoustic characteristics due to the biological aging process. Those acoustic characteristics, that are known to change throughout one’s life due to a physiological aging process, are used as variables in the age estimation rule. In Hollien’s model, speaker’s sex is treated as another variable because acoustic characteristics associated with aging are different and changes differently over time. In this model, a listener could have two estimation rules: one for male and one for female. However, the same rules would apply to a speaker of any language because the physiological model of vocal age is language-universal.

**Evaluations**

The four models described above will be assessed by examining results in the current study. First, the current study found that the listener’s age estimation was better when they judged the age of talkers in their native language than in the foreign language in the phrase and sentence conditions. The effect of language familiarity is not explained by the abstractionist model because the abstractionist model predicts that a listener’s age

\(^6\) Information about a talker’s age is considered as completely irrelevant information to descriptions of sound systems in classic linguistic theory (Jakobson, Fant, & Halle, 1963; Chomsky & Halle, 1968; Stevens & Blumstein, 1978).
estimation is not affected by a talker’s language. On the other hand, the effect of language familiarity can be explained in the other three models. All three models predict poor age estimation for foreign speech because relative to the number of stored exemplars for a listener’s own native language, there is insufficient number of stored examples (people) that the listener could access to in order to retrieve the age information for foreign speakers. Also, more variability is expected for foreign speakers in the prototype model due to less accurate mapping at the classification process.

All models could account for the effect of linguistic context (i.e., Vowel < Phrase < Sentence). It is considered that the number of perceptual cues that enable the listeners to identify the age of talker increases as the amount information in the stimuli increases. In the abstractionist model, it is presumed that there are some invariant features to determine the age of talker. Age estimation errors occur when some of the variables in the age estimation rule are not available in speech signal. Better age estimation in the sentence condition implies that there are more “age features” in the sentence level than in the vowel and phrase levels. In the other three models, the sentence stimuli add or multiply acoustic dimensions more than the phrase than the vowel stimuli, so that the listeners can identify the age of talkers more easily and accurately.

The abstractionist model does not account for the current result that age estimation of listeners was better for female than male talkers. The effect of talker’s sex can be attributed to the difference in the stored examples in the other three models. In these models, the sex difference could be due to qualitative or quantitative differences in

---

7 Foreign speech is equivalent to a less familiar language in this study. Predictions would be different if listener had an extensive knowledge or exposure to the language. We will not discuss cases like bilingual or multilingual listeners here.
exemplars of male and female talkers. In the prototype model, the sex difference could be also related to the classification process. If the prototypes have averaged properties of male and female talkers, perceptual similarity to the prototype is the same for male and female elderly talkers because the sex-related differences become less distinctive among the elderly talkers (e.g. Hollien, 1995); hence the difference between male and female talkers is not expected for elderly talkers. For the middle-aged talkers, the perceptual similarity between the prototype and middle-aged male talkers could be greater relative to the age-matched female talkers because age-related physiological changes appeared earlier in male than female. So, the predictions of the prototype model partially match with the current results. However, the prototype model does not predict a better estimation for female speech than male speech as long as the prototypes of female speech are equally good as the prototypes of male speech.

Whether perception of vocal age requires direct access to the listener’s stored information for specific speakers as in the exemplar model and the anchoring model, or age estimation is based on the prototypical speech for a certain age in the prototype model, the implication is that age-related speech characteristics are extracted based on individual speakers at some point. Thus, the accuracy of age estimation depends on the amount of linguistic experience that a listener has. The amount of speaker-specific information that the listener stored then determines how finely-tuned the scale of perceived age is. The scale of perceived age is usually not coarse but rather continuous (e.g. Shipp & Hollien, 1969), but some individual differences in the perception results could be explained by the difference in quantity of speaker information that each listener stored.
If age perception is based on exemplars of individual speakers as we proposed above, the process of perceptual estimation for talker’s age involves a search of similar talkers to the unknown talker in question. Based on the talker selected as the closest match, the age of a talker will be estimated. There is a problem in this process. The age of speakers can be inaccurately labeled and stored. If so, the retrieved age information is also inaccurate.

The anchoring model does not have a problem related with inaccurate age information because a listener determines the age of an unknown talker by comparing people for whom the listener knows their age. However, the anchoring model could have a limitation of its generalization. The anchoring model might not be robust enough to capture all mental images that the listener needs for particular persons at different points in their life. People change their voice throughout their life time\(^8\), and people sometimes remember those changes. The mental image for someone’s voice could be different from images for the same person’s voice in youth. For example, people remember how famous actors or newscasters sounded when they were young. Some of the middle-aged and elderly talkers told the experimenter during the recording, “I had a beautiful voice when I was young.” This dynamic of speaker information is not well-captured in the anchoring model. In the exemplar model, since all instances are valued and contributed to the final listener’s decision, it would not be a problem for a listener to access to the information of a certain speaker in different times.

This dissertation study employed chronological age as a reference measure. However, people do not have such an accurate age mapping from daily experiences.

\(^8\) Some of the individual speech characteristics may remain similar through his/her life.
Inaccurate age information can provide explanations as to why the listeners tended to underestimate the age for middle-aged and elderly talkers, and moreover why the Japanese listeners showed a better age estimation performance than the English listeners, as will be explored in the next section.

**Accessibility of age information**

There is no doubt that age plays an important role in our daily life whether it is in the United States or in Japan. However, there is a difference between the two countries in terms of the degree of accessibility to age information. Compared to Japanese society, it is difficult to find out a person’s exact age in the United States. Except family members, close relatives or friends, people only estimate an approximate age of a person based on their appearance, or personal information such as jobs, friends, or family members. Hence, the age property for a certain speaker might not be so neatly and carefully catalogued by the American people. On the other hand, knowing the exact age of individual persons is critical to establish smooth communications for virtually any occasion in Japan.

Japanese is a language that uses honorific words and expressions\(^9\). The appropriate usage of Japanese honorific expressions is determined by the distance of social position (or rank) and the closeness of personal relationship among talkers. Of course, age is not the only factor to determine the social and personal relationship between talkers, but Japanese social structure is still constructed principally by a person’s age. Rank of position at work depends on the seniority system in general. Overall it

\(^9\) People have been alarmed that honorific forms are now less frequent, and used wrongly more often in the contemporary Japanese society.
appears that there are more opportunities for Japanese to figure out the exact age of a person. Thus, in Japan, information about the age of speakers could be neatly arrayed in chronological order; and furthermore, each talker could be cross-referenced with each other. This means that Japanese may tend to have a larger database of stored speech representations attached to a wider range of ages, where the age information is more accurately mapped to a specific speaker.

**Intergenerational social interactions**

Another reason for the underestimation of elderly talkers by the young listeners could be related with the social network that the younger people are actively participated in. Although chronological age is a continuous scale, there is a social discontinuity between the young and the elderly people. There has become less of a number of multigenerational social networks available in the United States as well as in Japan, and less interaction between age-oriented social networks. Note that the retirement system, especially in the United States, creates increased socio-cultural separation between the generations. The proportion of three generation households in the United States also has been low relative to Japan. In the interview after the perception experiment, the author became aware of how remotely the young English listeners view the people in their seventies or eighties. For a majority of the young American listeners, people in the fifties or sixties were “old people”, which matches the result that the mean perceived age for the English elderly talkers by English listeners (in phrase and sentence contexts) was 60 years. On the contrary, the mean perceived age for the Japanese elderly by Japanese listeners (in the phrase and sentence contexts) was 71 years. This is within the age of most grandparents whom majority of the young Japanese listeners in this study lived with.
Presumably, such psychological distance to the elderly people from the young English listeners could compress the scale used for perceived age. The consequence of compressed age scaling can provide an explanation for the results of relatively large sizes of age estimation errors despite the high correlations between the chronological age and the perceived age (in phrase and sentence conditions). Remoteness to the elderly generations seems to be stronger among the English listeners than the Japanese in the current study. Such a social difference in recognition toward the elderly could be a reason why the Japanese listeners exhibited better performance on age estimation.

**Language specific acoustic characteristics**

A cross-language comparison for acoustic characteristics has some limitation because no matter how similar the linguistic and phonetic contexts of speech materials are between the two languages, there is no exact correspondence. It is difficult to conduct cross-language acoustic analysis of the sentence stimuli because phonetic contexts differ in the two languages. The current study employed the phrase stimuli in order to reduce contextual differences between the languages. The letter sequence used in the phrase stimuli ‘BCC’ has a very similar segmental structure in both languages. However, as we discussed in Chapter 3, there are slight differences in the pronunciation of each segment between Japanese and English (e.g. Voice onset time for /b/; spectral differences in the consonant pronunciation for the letter C).

In addition to these differences, there could be a difference in prosody; specifically the accentual patterns of the phrase stimuli are different between English talkers and Japanese talkers. Some talkers pronounced each letter as a separate accentual phrase, while others grouped the three syllables into a single phrase. The experimenter
noted that separate phrases happened more frequently in Japanese elderly talkers than other groups of talkers. This was also observed when Japanese younger talkers imitated speech of the elderly talkers. In these utterances in Japanese, there is a pitch fall within each syllable. **Figure 18** shows examples of pitch contours of the phrase *BCC* produced by the middle-aged English woman (Top) and the middle-aged Japanese woman (Bottom). As can be seen, Japanese speech shows a pitch declination in addition to the distinctive pitch fall in the syllable\(^\text{10}\). These pitch behavior in the Japanese phrase stimuli could be perceived distinctively odd to English listeners. However, such a dynamic difference in prosodic characteristics will not be captured if we compare average fundamental frequency or range of fundamental frequency. Whether segmental or prosodic, the acoustic differences introduced by language specific characteristics could influence the listener’s estimation of talker’s age. Detailed acoustic analysis will deepen our understanding how listeners perceive certain acoustic characteristics to identify speech and/or voice properties of individual speakers.

\(^\text{10}\) It may be not generalized to all Japanese speakers. Pitch fall within monosyllabic word is known as distinctive characteristics of Kansai accent. Japanese talkers in this study were all speakers of some variety of Kansai dialect.
Figure 18. Pitch contours of the phrase stimulus by English middle-aged woman, EFM07 (Top panel) and Japanese middle-aged woman, JFM04 (Bottom panel) in the normal condition.

In addition to the speech differences derived from an individual linguistic system, sociolinguistic and para-linguistic differences exist in the two languages in question. Although the perceptual results of age-disguised speech did not exhibit the effect of language, it seems that there is a cross-language difference in the perceptual speech
characteristics associated with a certain age group. Remember that this study collected speech samples in which each talker impersonated five types of speakers, i.e. baby, mother, father, grandmother, and grandfather. The author noticed an interesting difference in the way talkers tried to impersonate a grandmother (bear). English talkers tended to employ a high pitch when they imitated a grandmother, whereas Japanese talkers tended to speak at their lower pitch when they imitated a grandmother. This suggests that the speech characteristics typically associated to certain age groups are different in the United States and Japan. Thus, it is possible that different perceptual stereotypes associated to a certain talker group affected the listener’s judgment on talker’s age. For example, stereotypes may provide an explanation as to why English females were perceived particularly young to Japanese listeners. Acoustic analysis of the age-disguised and impersonated speech will be made and is expected that the detailed acoustic results will clarify the relationship between the perceived age and language-specific characteristics.

Future research

Following the current dissertation study, it would be interesting to extend the same paradigm to other languages in order to examine the perceptual effects of language-specific and language-universal speech characteristics. For example, in order to clarify issues of language familiarity on age perception, it would be interesting to conduct a cross-language study with an extremely unfamiliar language compared with English and Japanese. There are many potential languages, but one of the North American indigenous languages would be particularly interesting. Languages such as Lacota, Navajo, Mohawk, Muskogee, Koasati, and Chickasaw, are spoken by relatively large
number of native speakers among the indigenous languages in the North America. Except a few people, these languages would be extremely unfamiliar to both English and Japanese listeners. Despite the fact that existing native speakers are all elderly speakers in many endangered languages, aging has been a variable that has not been closely examined by linguists working on the endangered languages. It is expected that the proposed research could provide the referential data on speech characteristics associated with aging common to languages, and furthermore contribute to accurate documentation of endangered languages.

As mentioned in chapter 1, acoustic correlates to perceived age have been examined by various researchers. However, it is yet unclear whether those findings are confined to a specific language or to what extent they can be applied to another language. For example, changes in voice quality are considered the most distinctive perceptual property associated to elderly speakers. Noise components in the voice source and irregularity in F0 are used as acoustic measures of voice quality, such as breathiness, harshness, or roughness; and those acoustic properties are known to be highly correlated with listener’s perceived age of talkers (e.g. Amerman & Parnell, 1990; Ryan & Burk, 1974; Braun & Rietveld, 1995). However, in many languages, voice quality serves as linguistically relevant information. Breathy sounds are distinguished from modal sounds as linguistically contrastive sounds in many languages such as Hindi, Shindi, Tsonga, and Gujarati. Laryngealized vowels or creaky voices are found in Mazatec, Bruu, and Bura as a part of their phonetic inventories. It was also reported that voice quality marks phrase boundaries in Swedish (Fant & Kruckenberg, 1989). Hence, it would be interesting to
examine the language effect on age perception in a language where voice qualities are used to distinguish phonetic sounds in the phonological system.

Pitch is another perceptual cue that has been extensively studied to find the acoustic correlates with perceived age. High correlations were found between perceived age and standard deviation of speaking fundamental frequency, and also between perceived age and mean fundamental frequency. Pitch, however, carries various information in speech. In addition to carrying the paralinguistic information such as attitudes and emotion, pitch is used for differentiating tonal distinctions, signaling accents and phrasal boundaries, and emphasizing a specific word. Pitch contours within a phrase or a sentence would be quite different across languages, as we have seen the comparison of pitch contours between the two languages above in Figure 18. It would be quite interesting to know how these language-oriented differences in pitch behavior could influence the listener’s perception of speaker characteristics.

A language that uses contour tones as their phonological features such as Mandarin, Vietnamese, and Thai, would provide valuable information on the effect of pitch change on age perception. Since pitch movements within a syllable are not common in English utterances, it could be judged as disordered speech in the diagnosis for English speakers. Since there could be more than three generations from these language speakers, the results could be extended to the study of non-native elderly speakers in the United States.

The results of multi-language comparison would enable us to distinguish speech characteristics changes in lifespan inherent in biological aging from speech
characteristics caused by historical and environmental changes in a particular speech community or group.

**Conclusion**

The main aim of this research was to investigate the perceptual effects of linguistic and para-linguistic factors on the listener’s estimation of talker’s age. In line with previous findings, moderate to high correlations between talker’s chronological age and listener’s perceived age were observed for a familiar native language. Similar correlations were also found for a less familiar foreign language. Regardless of the language familiarity, the listeners’ age estimation became better as the amount of information in the stimuli increased. More importantly, the results provided clear evidence that language familiarity affects age perception. Also, the age of female talkers is easier to estimate than that of male talkers, but only by listeners who know the language of the talker. Furthermore, Japanese listeners identify the age of speakers more accurately than English listeners. These results suggest that age-related speech characteristics are based on not only purely physiological factors but are also related to specifically linguistic variation, variation that a non-native listener does not have access to.

Finally, results showed no significant difference in terms of listener’s language in the age perception for age-disguised speech. These results imply that the stereotypical speech images associated to younger or older talkers may be similar across different language communities. However, results of current study also imply that a listener’s linguistic background can affect how they perceive indexical properties of speech such as age and sex. It would be beneficial in speech-language clinics for non-native patients if
speech language pathologies were aware that listener’s experiences with a given language could affect in the assessment of speech and voice disorders for patients with linguistic backgrounds different from clinician’s.
References


Appendix A1: Information of English talkers. If the person was a student, the degree pursued at the time of recording was listed.

<table>
<thead>
<tr>
<th>ID</th>
<th>Age</th>
<th>Sex</th>
<th>Talker original locations</th>
<th>Foreign language(s) spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td>efy01</td>
<td>25</td>
<td>F</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>efy02</td>
<td>25</td>
<td>F</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>efy04</td>
<td>28</td>
<td>F</td>
<td>OH</td>
<td>-</td>
</tr>
<tr>
<td>efy05</td>
<td>24</td>
<td>F</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>efy06</td>
<td>25</td>
<td>F</td>
<td>OH</td>
<td>Spanish</td>
</tr>
<tr>
<td>efm01</td>
<td>57</td>
<td>F</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>efm02</td>
<td>56</td>
<td>F</td>
<td>OH</td>
<td>-</td>
</tr>
<tr>
<td>efm03</td>
<td>56</td>
<td>F</td>
<td>ME</td>
<td>French</td>
</tr>
<tr>
<td>efm06</td>
<td>60</td>
<td>F</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>efm07</td>
<td>60</td>
<td>F</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>efo01</td>
<td>80</td>
<td>F</td>
<td>LA</td>
<td>-</td>
</tr>
<tr>
<td>efo02</td>
<td>80</td>
<td>F</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>efo04</td>
<td>86</td>
<td>F</td>
<td>PA</td>
<td>-</td>
</tr>
<tr>
<td>efo05</td>
<td>81</td>
<td>F</td>
<td>IL</td>
<td>-</td>
</tr>
<tr>
<td>efo06</td>
<td>84</td>
<td>F</td>
<td>PA</td>
<td>-</td>
</tr>
<tr>
<td>emy02</td>
<td>30</td>
<td>M</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>emy04</td>
<td>25</td>
<td>M</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>emy05</td>
<td>26</td>
<td>M</td>
<td>IN</td>
<td>Spanish, Chinese</td>
</tr>
<tr>
<td>emy07</td>
<td>26</td>
<td>M</td>
<td>IN</td>
<td>Spanish</td>
</tr>
<tr>
<td>emy08</td>
<td>24</td>
<td>M</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>emm01</td>
<td>55</td>
<td>M</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>emm02</td>
<td>58</td>
<td>M</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>emm03</td>
<td>59</td>
<td>M</td>
<td>IN</td>
<td>French</td>
</tr>
<tr>
<td>emm04</td>
<td>55</td>
<td>M</td>
<td>IN</td>
<td>Spanish, Portuguese, Nahuatl</td>
</tr>
<tr>
<td>emm06</td>
<td>54</td>
<td>M</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>emo01</td>
<td>82</td>
<td>M</td>
<td>IN</td>
<td>French, German</td>
</tr>
<tr>
<td>emo02</td>
<td>82</td>
<td>M</td>
<td>IA</td>
<td>-</td>
</tr>
<tr>
<td>emo03</td>
<td>83</td>
<td>M</td>
<td>IN</td>
<td>-</td>
</tr>
<tr>
<td>emo04</td>
<td>80</td>
<td>M</td>
<td>IN</td>
<td>French, German, Italian</td>
</tr>
<tr>
<td>emo05</td>
<td>83</td>
<td>M</td>
<td>AZ</td>
<td>French, Dutch</td>
</tr>
</tbody>
</table>
## Appendix A1: Information of English talkers (continue).

<table>
<thead>
<tr>
<th>ID</th>
<th>Occupation</th>
<th>Education</th>
<th>Professional voice training</th>
<th>Cohabiters</th>
</tr>
</thead>
<tbody>
<tr>
<td>efy01</td>
<td>student</td>
<td>Master</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>efy02</td>
<td>SLP</td>
<td>Master</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>efy04</td>
<td>high school teacher</td>
<td>Bachelor</td>
<td>none</td>
<td>S, C</td>
</tr>
<tr>
<td>efy05</td>
<td>housewife</td>
<td>Associates</td>
<td>none</td>
<td>S, C</td>
</tr>
<tr>
<td>efy06</td>
<td>community living specialist</td>
<td>Bachelor</td>
<td>none</td>
<td>F</td>
</tr>
<tr>
<td>efm01</td>
<td>elementary school principal</td>
<td>Master</td>
<td>none</td>
<td>S, C</td>
</tr>
<tr>
<td>efm02</td>
<td>food site manager</td>
<td>Diploma</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>efm03</td>
<td>marketing director</td>
<td>Master</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>efm06</td>
<td>retired (elementary teacher)</td>
<td>Master</td>
<td>none</td>
<td>S, C</td>
</tr>
<tr>
<td>efm07</td>
<td>administrative assistant</td>
<td>Associates</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>efo01</td>
<td>retired (school teacher)</td>
<td>Bachelor</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>efo02</td>
<td>retired (secretary)</td>
<td>Bachelor</td>
<td>singing</td>
<td>none</td>
</tr>
<tr>
<td>efo04</td>
<td>retired (professional house work)</td>
<td>Diploma</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>efo05</td>
<td>retired (book keeper)</td>
<td>Diploma</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>efo06</td>
<td>retired (public relations)</td>
<td>Diploma</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>emy02</td>
<td>student</td>
<td>Doctoral</td>
<td>singing</td>
<td>S, C</td>
</tr>
<tr>
<td>emy04</td>
<td>student</td>
<td>Doctoral</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>emy05</td>
<td>student</td>
<td>Bachelor</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>emy07</td>
<td>student</td>
<td>Master</td>
<td>none</td>
<td>F, O</td>
</tr>
<tr>
<td>emy08</td>
<td>student</td>
<td>Master</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>emm01</td>
<td>computer programmer</td>
<td>Associates</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>emm02</td>
<td>computer programmer</td>
<td>Master</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>emm03</td>
<td>lawyer</td>
<td>Doctoral</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>emm04</td>
<td>computer programmer</td>
<td>Master</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>emm06</td>
<td>potter/blacksmith</td>
<td>Bachelor</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>emo01</td>
<td>retired (university administrator)</td>
<td>Master</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>emo02</td>
<td>retired (accountant)</td>
<td>Bachelor</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>emo03</td>
<td>retired (insurance agent)</td>
<td>Bachelor</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>emo04</td>
<td>retired (teacher)</td>
<td>Master</td>
<td>singing, acting</td>
<td>S</td>
</tr>
<tr>
<td>emo05</td>
<td>retired (professor)</td>
<td>Doctoral</td>
<td>none</td>
<td>S</td>
</tr>
</tbody>
</table>

S: Spouse, C: Child(ren), F: Friend(s), O: Other
### Appendix A1: Information of English talkers (continue).

<table>
<thead>
<tr>
<th>ID</th>
<th>Present smoking habit</th>
<th>Years of smoking</th>
<th>Years after quitting smoking</th>
<th>Current physical activity level</th>
<th>Teeth condition</th>
<th>Present physical status</th>
</tr>
</thead>
<tbody>
<tr>
<td>efy01</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>efy02</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>efy04</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>efy05</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>efy06</td>
<td>no</td>
<td>6</td>
<td>2 mon.</td>
<td>moderate</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>efm01</td>
<td>no</td>
<td>8</td>
<td>30</td>
<td>moderate</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>efm02</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>Upper denture</td>
<td>CHD, HBP, CB, asthma</td>
</tr>
<tr>
<td>efm03</td>
<td>no</td>
<td>6</td>
<td>30</td>
<td>mild</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>efm06</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>efm07</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>good</td>
<td>HBP</td>
</tr>
<tr>
<td>efo01</td>
<td>no</td>
<td>10</td>
<td>20</td>
<td>mild</td>
<td>good</td>
<td>CHD, HA, HBP, CB</td>
</tr>
<tr>
<td>efo02</td>
<td>no</td>
<td>15</td>
<td>40</td>
<td>mild</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>efo04</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>fair</td>
<td>HBP</td>
</tr>
<tr>
<td>efo05</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>fair</td>
<td>CHD, HBP</td>
</tr>
<tr>
<td>efo06</td>
<td>no</td>
<td>20</td>
<td>40</td>
<td>mild</td>
<td>fair</td>
<td>HA</td>
</tr>
<tr>
<td>emy02</td>
<td>no</td>
<td>9</td>
<td>1</td>
<td>mild</td>
<td>good</td>
<td>CR</td>
</tr>
<tr>
<td>emy04</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>emy05</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>emy07</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>high</td>
<td>good</td>
<td>HBP</td>
</tr>
<tr>
<td>emy08</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>emm01</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>high</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>emm02</td>
<td>no</td>
<td>20</td>
<td>10</td>
<td>mild</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>emm03</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>high</td>
<td>good</td>
<td>-</td>
</tr>
<tr>
<td>emm04</td>
<td>no</td>
<td>30</td>
<td>8</td>
<td>moderate</td>
<td>good</td>
<td>HBP</td>
</tr>
<tr>
<td>emm06</td>
<td>yes</td>
<td>29</td>
<td>0</td>
<td>moderate</td>
<td>good</td>
<td>HBP</td>
</tr>
<tr>
<td>emo01</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>fair</td>
<td>HA</td>
</tr>
<tr>
<td>emo02</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>good</td>
<td>HBP</td>
</tr>
<tr>
<td>emo03</td>
<td>no</td>
<td>15</td>
<td>35</td>
<td>mild</td>
<td>fair</td>
<td>-</td>
</tr>
<tr>
<td>emo04</td>
<td>no</td>
<td>19</td>
<td>42</td>
<td>high</td>
<td>good</td>
<td>CHD</td>
</tr>
<tr>
<td>emo05</td>
<td>no</td>
<td>15</td>
<td>40</td>
<td>moderate</td>
<td>good</td>
<td>-</td>
</tr>
</tbody>
</table>

HBP: High blood pressure, CHD: Coronary heart disease, HA: Heart arrhythmia, CB: Chronic bronchitis, CR: Chronic rhinitis
Appendix A2: Information of Japanese talkers. If the person was a student, the degree pursued at the time of recording was listed.

<table>
<thead>
<tr>
<th>ID</th>
<th>CA</th>
<th>Sex</th>
<th>Talkers original location</th>
<th>Foreign language(s) spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td>jfy02</td>
<td>28</td>
<td>F</td>
<td>Osaka</td>
<td>-</td>
</tr>
<tr>
<td>jfy03</td>
<td>28</td>
<td>F</td>
<td>Hyogo*</td>
<td>-</td>
</tr>
<tr>
<td>jfy05</td>
<td>29</td>
<td>F</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jfy06</td>
<td>25</td>
<td>F</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jfy08</td>
<td>25</td>
<td>F</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jfm01</td>
<td>55</td>
<td>F</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jfm04</td>
<td>58</td>
<td>F</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jfm05</td>
<td>58</td>
<td>F</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jfm06</td>
<td>56</td>
<td>F</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jfm07</td>
<td>57</td>
<td>F</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jfo02</td>
<td>80</td>
<td>F</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jfo03</td>
<td>86</td>
<td>F</td>
<td>Kochi</td>
<td>-</td>
</tr>
<tr>
<td>jfo04</td>
<td>80</td>
<td>F</td>
<td>Wakayama</td>
<td>-</td>
</tr>
<tr>
<td>jfo07</td>
<td>84</td>
<td>F</td>
<td>Osaka</td>
<td>-</td>
</tr>
<tr>
<td>jfo08</td>
<td>81</td>
<td>F</td>
<td>Kobe</td>
<td>-</td>
</tr>
<tr>
<td>jmy01</td>
<td>30</td>
<td>M</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jmy02</td>
<td>26</td>
<td>M</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jmy03</td>
<td>30</td>
<td>M</td>
<td>Kyoto</td>
<td>-</td>
</tr>
<tr>
<td>jmy04</td>
<td>28</td>
<td>M</td>
<td>Osaka</td>
<td>-</td>
</tr>
<tr>
<td>jmy06</td>
<td>27</td>
<td>M</td>
<td>Osaka</td>
<td>-</td>
</tr>
<tr>
<td>jmm01</td>
<td>55</td>
<td>M</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jmm02</td>
<td>55</td>
<td>M</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jmm03</td>
<td>57</td>
<td>M</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jmm04</td>
<td>56</td>
<td>M</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jmm05</td>
<td>60</td>
<td>M</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jmo01</td>
<td>86</td>
<td>M</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jmo02</td>
<td>80</td>
<td>M</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jmo03</td>
<td>83</td>
<td>M</td>
<td>Hyogo</td>
<td>-</td>
</tr>
<tr>
<td>jmo05</td>
<td>80</td>
<td>M</td>
<td>Osaka</td>
<td>-</td>
</tr>
<tr>
<td>jmo06</td>
<td>85</td>
<td>M</td>
<td>Osaka</td>
<td>-</td>
</tr>
</tbody>
</table>

* All talkers originally from Hyogo were from the southern east parts of Hyogo prefecture.
Appendix A2: Information of Japanese talkers (continue).

<table>
<thead>
<tr>
<th>ID</th>
<th>Occupation</th>
<th>Education</th>
<th>Professional voice training</th>
<th>Cohabiters</th>
</tr>
</thead>
<tbody>
<tr>
<td>jfy02</td>
<td>student</td>
<td>grad. school</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>jfy03</td>
<td>student</td>
<td>grad. school</td>
<td>none</td>
<td>M, F, Si, R</td>
</tr>
<tr>
<td>jfy05</td>
<td>student</td>
<td>grad. school</td>
<td>none</td>
<td>M, F</td>
</tr>
<tr>
<td>jfy06</td>
<td>student</td>
<td>grad. school</td>
<td>none</td>
<td>F</td>
</tr>
<tr>
<td>jfy08</td>
<td>office worker</td>
<td>university</td>
<td>none</td>
<td>M, F, Si</td>
</tr>
<tr>
<td>jfm01</td>
<td>housewife (former teacher)</td>
<td>university</td>
<td>none</td>
<td>S, C, M</td>
</tr>
<tr>
<td>jfm04</td>
<td>retired (office worker)</td>
<td>high school</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>jfm05</td>
<td>housewife</td>
<td>high school</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>jfm06</td>
<td>housewife</td>
<td>college</td>
<td>none</td>
<td>S, C, M</td>
</tr>
<tr>
<td>jfm07</td>
<td>housewife</td>
<td>high school</td>
<td>none</td>
<td>S, C, Gc, CS</td>
</tr>
<tr>
<td>jfo02</td>
<td>retired (housewife)</td>
<td>girl's school</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>jfo03</td>
<td>retired (housewife)</td>
<td>normal school</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>jfo04</td>
<td>retired (rice store management)</td>
<td>girl's school</td>
<td>none</td>
<td>S, C, Gc, CS</td>
</tr>
<tr>
<td>jfo07</td>
<td>retired (part-time public officer)</td>
<td>girl's school</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>jfo08</td>
<td>landlord</td>
<td>professional school</td>
<td>singing</td>
<td>none</td>
</tr>
<tr>
<td>jmy01</td>
<td>teacher</td>
<td>grad. school</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>jmy02</td>
<td>student</td>
<td>grad. school</td>
<td>none</td>
<td>M, F, Si</td>
</tr>
<tr>
<td>jmy03</td>
<td>student</td>
<td>grad. school</td>
<td>none</td>
<td>M, F, Si</td>
</tr>
<tr>
<td>jmy04</td>
<td>student</td>
<td>grad. school</td>
<td>none</td>
<td>M, F</td>
</tr>
<tr>
<td>jmy06</td>
<td>public officer</td>
<td>university</td>
<td>none</td>
<td>M, F, Si</td>
</tr>
<tr>
<td>jmm01</td>
<td>office worker</td>
<td>university</td>
<td>none</td>
<td>S, C</td>
</tr>
<tr>
<td>jmm02</td>
<td>real estate agent</td>
<td>grad. school</td>
<td>none</td>
<td>S, C</td>
</tr>
<tr>
<td>jmm03</td>
<td>office worker</td>
<td>college</td>
<td>none</td>
<td>M</td>
</tr>
<tr>
<td>jmm04</td>
<td>news dealer</td>
<td>high school</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>jmm05</td>
<td>marketing consultant</td>
<td>university</td>
<td>none</td>
<td>S, C</td>
</tr>
<tr>
<td>jmo01</td>
<td>retired (telephone service)</td>
<td>grad. school</td>
<td>none</td>
<td>S, C, CS</td>
</tr>
<tr>
<td>jmo02</td>
<td>retired (professor)</td>
<td>grad. school</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>jmo03</td>
<td>retired (office worker)</td>
<td>grad. school</td>
<td>none</td>
<td>S</td>
</tr>
<tr>
<td>jmo05</td>
<td>retired (office worker)</td>
<td>grad. school</td>
<td>singing</td>
<td>S</td>
</tr>
<tr>
<td>jmo06</td>
<td>parking management</td>
<td>university</td>
<td>singing</td>
<td>none</td>
</tr>
</tbody>
</table>

S: Spouse, C: Child(ren), M: Mother, F: Father, Gc: Grandchild(ren), Gm: grandmother, Gf: Grandfather, CS: Child(ren)’s spouse, Si: Siblings, R: Other relatives, F: Friend(s),
## Appendix A2: Information of Japanese talkers (continue).

<table>
<thead>
<tr>
<th>ID</th>
<th>Present smoking habit</th>
<th>Years of smoking</th>
<th>Years after quitting smoking</th>
<th>Current physical activity level</th>
<th>Teeth condition</th>
<th>Present physical status</th>
</tr>
</thead>
<tbody>
<tr>
<td>jfy02</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jfy03</td>
<td>yes</td>
<td>7</td>
<td>0</td>
<td>mild</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jfy05</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jfy06</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jfy08</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>good</td>
<td>CB</td>
</tr>
<tr>
<td>jfm01</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>fair</td>
<td>CR, adenoidectomy</td>
</tr>
<tr>
<td>jfm04</td>
<td>no</td>
<td>10</td>
<td>8</td>
<td>moderate</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jfm05</td>
<td>no</td>
<td>20</td>
<td>3</td>
<td>mild</td>
<td>good</td>
<td>CR</td>
</tr>
<tr>
<td>jfm06</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>fair</td>
<td>HBP</td>
</tr>
<tr>
<td>jfm07</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>good</td>
<td>CR, CED</td>
</tr>
<tr>
<td>jfo02</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>upper denture</td>
<td>no</td>
</tr>
<tr>
<td>jfo03</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>fair</td>
<td>CED</td>
</tr>
<tr>
<td>jfo04</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>upper denture</td>
<td>HBP</td>
</tr>
<tr>
<td>jfo07</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jfo08</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>lower denture</td>
<td>HBP</td>
</tr>
<tr>
<td>jmy01</td>
<td>no</td>
<td>10</td>
<td>1</td>
<td>mild</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jmy02</td>
<td>yes</td>
<td>9</td>
<td>0</td>
<td>mild</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jmy03</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jmy04</td>
<td>no</td>
<td>6</td>
<td>2</td>
<td>mild</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jmy06</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>good</td>
<td>no</td>
</tr>
<tr>
<td>jmm01</td>
<td>no</td>
<td>10</td>
<td>15</td>
<td>mild</td>
<td>good</td>
<td>CR</td>
</tr>
<tr>
<td>jmm02</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>mild</td>
<td>good</td>
<td>CR</td>
</tr>
<tr>
<td>jmm03</td>
<td>no</td>
<td>12</td>
<td>27</td>
<td>mild</td>
<td>good</td>
<td>HBP, CR, HA</td>
</tr>
<tr>
<td>jmm04</td>
<td>yes</td>
<td>35</td>
<td>2 y w/o smoking</td>
<td>high</td>
<td>upper denture</td>
<td>adenoidectomy</td>
</tr>
<tr>
<td>jmm05</td>
<td>yes</td>
<td>40</td>
<td>na</td>
<td>moderate</td>
<td>upper denture</td>
<td>adenoidectomy</td>
</tr>
<tr>
<td>jmo01</td>
<td>no</td>
<td>0</td>
<td>na</td>
<td>moderate</td>
<td>dentures (U &amp; L)</td>
<td>HBP</td>
</tr>
<tr>
<td>jmo02</td>
<td>no</td>
<td>40</td>
<td>20</td>
<td>mild</td>
<td>dentures (U &amp; L)</td>
<td>HA, HBP, CR</td>
</tr>
<tr>
<td>jmo03</td>
<td>no</td>
<td>60</td>
<td>3</td>
<td>mild</td>
<td>upper denture</td>
<td>HA, HBP, CED</td>
</tr>
<tr>
<td>jmo05</td>
<td>no</td>
<td>50</td>
<td>3</td>
<td>moderate</td>
<td>upper denture</td>
<td>HA, HBP, CED</td>
</tr>
<tr>
<td>jmo06</td>
<td>no</td>
<td>20</td>
<td>40</td>
<td>mild</td>
<td>dentures (U &amp; L)</td>
<td>no</td>
</tr>
</tbody>
</table>

HBP: High blood pressure, CHD: Coronary heart disease, HA: Heart arrhythmia, CB: Chronic bronchitis, CR: Chronic rhinitis, CED: Chronic ear diseases.
Appendix B1: Questionnaire for a talker (English version). The actual response sheets were written in a 14-point font size.

Please answer the following questions. Please ask me if you would like any assistance in filling the questionnaire out.

A. Background Information
(1) Sex: Male ____ Female ____
(2) Ethnicity:
   a) Hispanic or Latino__  b) Not Hispanic or Latino __
(3) Race:
   a) American Indian or Alaska Native __
   b) Asian ___
   c) Black or African American ___
   d) Native Hawaiian or other Pacific Islander ___
   e) White__
(4) Date of birth: Month___ Day___Year 19____
(5) What is your occupation?
   (1) ______________________
   (2) If retired, what was your occupation? ______________________
(6) What is your native language? ________________________________
(7) Do you know any other language(s)?
   (1) Yes____ No____
   (2) If yes, please list the language(s) you know.
   __________________________________________________________
(8) Do you use any other language(s) on a regular basis?
   (1) Yes____ No____
   (2) If yes, please list the language(s) you use.
   ______________________________
(9) Where do you come from?  City_________________________ State____________
(10) Please list the place(s) you have lived more than one year in chronological order.

<table>
<thead>
<tr>
<th>City name, State</th>
<th>Duration (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td></td>
</tr>
<tr>
<td>b)</td>
<td></td>
</tr>
<tr>
<td>c)</td>
<td></td>
</tr>
<tr>
<td>d)</td>
<td></td>
</tr>
<tr>
<td>e)</td>
<td></td>
</tr>
</tbody>
</table>
(11) Have you had professional voice training?
   (1) Yes____ No____
   (2) If yes, what was the training for? (Please check all that apply.)
       a) Public speaking ____
       b) Singing ____
       c) Acting ____
(12) Do you live with somebody else?
(1) Yes____ No____
(2) If yes, whom do you live with? (Please check all that apply.)
   a) Spouse ___
   b) Child(ren) ___
   c) Mother ___
   d) Father ___
   e) Child(ren)'s spouse ___
   f) Grandfather___
   g) Grandmother___
   h) Grandchild(ren)___
   i) Brother(s)/Sister(s)___
   j) Other family members___
   k) Friend(s)___
   l) Other ___

(13) What is your highest education level completed (or currently working on)?
   a) Less than high school ___
   b) High school Diploma ___
   c) Associates Degree ___
   d) Bachelor’s Degree ___
   e) Masters Degree ___
   f) Doctoral Degree ___
   g) Professional Degree ___
   h) Other:______________
B. Physical Activities

(1) Thinking about the things you do at work, how would you rate yourself as to the amount of physical activity you get compared with others of your age and sex? (Please circle one.)
   a) Much more active
   b) Somewhat more active
   c) About the same
   d) Somewhat less active
   e) Much less active
   f) Not applicable

(2) Thinking about the things you do at outside of work, how would you rate yourself as to the amount of physical activity you get compared with others of your age and sex? (Please circle one.)
   a) Much more active
   b) Somewhat more active
   c) About the same
   d) Somewhat less active
   e) Much less active
   f) Not applicable

(3) Please respond this question if you are age of 31 or above. If you are not, please go to the question (4).

Thinking about the things you did in the past at work and outside of work, how would you rate yourself as to the amount of physical activity you used to get compared with others of your age and sex? Please respond with using the following number for each decade of your age.

   a) Much more active
   b) Somewhat more active
   c) About the same
   d) Somewhat less active
   e) Much less active

   (1) When you were in your 20, ____
   (2) When you were in your 30, ____
   (3) When you were in your 40, ____
   (4) When you were in your 50, ____
   (5) When you were in your 60, ____
   (6) When you were in your 70, ____

(4) Do you regularly engage in strenuous exercise or hard physical labor?
   (1) Yes ___ No ___
   (2) If Yes, do you exercise or labor at least three times a week? Yes ___ No ___
(5) How often do you engage in vigorous-intensity level activities (digging in the garden, strenuous sports, jogging, aerobic dancing, swimming, brisk walking, heavy carpentry, bicycling on hills, etc.)? (Please circle one.)
   a) Once a day  ____
   b) every other day  ____
   c) once a week  ____
   d) (almost) never  ____

(6) How often do you engage in moderate-intensity level activities (housework, light sports, regular walking, golf, yard work, lawn mowing, wall painting, repairing, light carpentry, ballroom dancing, bicycling on level ground, etc.)? (Please circle one.)
   a) Once a day  ____
   b) every other day  ____
   c) once a week  ____
   d) (almost) never  ____

(7) Do you have difficulty to do any of the following activities by yourself? (Please check all that apply.)
   a) Bathing  ____
   b) Dressing  ____
   c) Walking  ____
   d) Eating  ____
   e) Toilet use  ____

(8) How many minutes do you regularly walk each day? _______ minutes/day

(9) How do you describe your usual pace of walking? (Please circle one.)
   a) Casual or strolling (less than 2mph)
   b) Average or normal (2-3 mph)
   c) Fairly brisk (3-4 mph)
   d) Brisk or striding (4 mph or faster)

(10) How many stairs do you climb up each day? ______ steps/day (Let 1 flight = 10 steps)

C. Dietary and Social Habits
(1) Do you smoke cigarettes?
   (1) Yes ____ No ____
   (2) If Yes, how many on an average day? _______
   (3) Age started? _______

(2) Did you ever smoke cigarettes?
   (1) Yes ____ No ____
   (2) If Yes, How many years have passed since you stopped smoking? ______
   (3) How many years did you smoke before you stopped smoking? ______
(3) How many servings of the following do you drink? Please check one for each category.

<table>
<thead>
<tr>
<th>Unit: Glass or cup</th>
<th>Almost never</th>
<th>1-2 per month</th>
<th>1-2 per week</th>
<th>3-6 per week</th>
<th>1-2 per day</th>
<th>3-5 per day</th>
<th>6+ per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Soft drink with caffeine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Coke, Mountain Dew etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Coffee</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Tea (with caffeine)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Beer, ale, stout, etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Wine, sherry, sake, etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Liquor (whiskey, gin, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**D. Past and Present Health Status**

In this section of the survey, we’d like to ask you about your personal information. Remember, we do not want you to answer any questions if you are uncomfortable doing so. Information you will provide for the current research is collected and used only for academic purposes, and will not be shared by anybody other than the principle investigator.

(1) Have you ever been diagnosed with a speech problem? Yes ___ No ___
(2) Have you ever been diagnosed with a language problem? Yes ___ No ___
(3) Have you noticed difficulty with your speech? Yes ___ No ___
(4) Have you been often asked to repeat what you said? Yes ___ No ___
(5) Have you excessively used your voice (shouting or yelling) in your activities at work, or any recreational events? Yes ___ No ___
(6) In your recent activities at work, or any recreational events, did you excessively use your voice (shouting or yelling)? Yes ___ No ___
(7) Is there a family history of speech/language problem? Yes ___ No ___
(8) Have you ever been diagnosed with a hearing problem? Yes ___ No ___
(9) Have you noticed difficulty with your hearing? Yes ___ No ___
(10) Do you have a history of ear infection or any pain in your ears? Yes ___ No ___
(11) Is there a family history of hearing loss? Yes ___ No ___
(12) Do you have any history of exposure to noise in recreational activities, at work, or in the military? Yes ___ No ___
(13) Do you experience dizziness or ringing (tinnitus) in ears? Yes ___ No ___
(14) Did you have any speech therapy when you were younger? Yes ___ No ___
(15) Do you have any teeth missing?
   (1) Yes ___ No ___
   (2) If Yes, How many?
      a) Front Uppers ___
      b) Front Lowers ___
      c) Back ___
(16) Do you wear a bridge? Yes ____ No ____
(17) Do you wear dentures?
   (1) Yes ____ No ____
   (2) If Yes, where?
      a) Uppers ____
      b) Lowers ____
      c) Both ____
(18) Do you think your teeth have an irregular spacing? Yes ____ No ____
(19) How would you describe your present state of health? (Please circle one.)
   a) Poor
   b) Fair
   c) Good
   d) Very good
   e) Excellent
(20) Has a physician ever told you that you had any of the following? (Please check and give year of onset, if applicable.)
   a) Coronary heart disease Yes ___ No ___ Year ______
   b) Heart arrhythmia Yes ___ No ___ Year ______
   c) Stroke Yes ___ No ___ Year ______
   d) High blood pressure Yes ___ No ___ Year ______
   e) Chronic bronchitis Yes ___ No ___ Year ______
   f) Asthma Yes ___ No ___ Year ______
   g) Chronic rhinitis Yes ___ No ___ Year ______
   h) Chronic ear diseases Yes ___ No ___ Year ______
   i) Depression Yes ___ No ___ Year ______
   j) Other neurological disease(s) Yes ___ No ___ Year ______
(21) Do you have any surgery or injury of the laryngeal area? Yes ____ No ____
(22) Do you take any of the following? (Please check all if applicable.)
   a) Antihistamines (e.g. Allegra, Clarinex, Claritin, or Benadryl) ___
   b) Amphetamines (e.g. Adderall) ___
   c) Sedatives (e.g. Benadryl, Valium, Ambien, or Diazepam Intensol) ___
   d) Drugs that functions as antispasmodic agents
      (e.g. Barbidonna, Bellatal, or Donnatal) ___
   e) I don’t know the name(s) of medicines I’m taking ___

The next question (23) is only for a female subject.
(23) Do you have menstruation?
   (1) Yes ___ No ___
   (2) If yes, is menstrual period regular or irregular? Regular ___ Irregular ___
Appendix B2: Questionnaire for a talker (Japanese version). The actual response sheets were written in a 14-point font size.

以下の質問事項に回答して下さい。必要な場合は、実験担当者が記入をお手伝いしますので、お知らせ下さい。

A. 基本事項
(1) 性別：男性・女性
(2) 民族：
   a) ヒスパニック・ラテン系 b) ヒスパニック・ラテン系以外
(3) 人種：
   a) アメリカ・インディアン及びアラスカ原住民
   b) アジア系
   c) 黒人・アフリカ系
   d) ハワイ原住民または太平洋諸島系
   e) 白人
(4) 生年月日： 大正・昭和 ______年____月____日（ ）歳
(5) 職業：
   (1) 現在の御職業____________________________________
   (2) 引退されている場合、以前の御職業 ______________________

(6) あなたの母国語は何ですか。________________________________
(7) 日本語以外の外国語を流暢に話せますか。
   (1) (はい・いいえ)
   (2) 「はい」と回答された方、よく知っている外国語をすべて書いて下さい

(8) 日本語以外の外国語を日常使用していますか。
   (1) (はい・いいえ)
   (2) 「はい」と回答された方、日常使用している外国語をすべて書いて下さい。 ______________________________________

(9) あなたの方言は何ですか。
   (1) 方言（大阪弁・京都弁など）________________________________
   (2) あなたの方言を習得した地域は主にどこですか。
       __________________府県____________________________市町村
（10）あなたの話す方言以外を話す地域に、一年以上滞在したことがありますか
（はい・いいえ）

（1）「はい」と回答された方にお聞きします。以前に一年以上滞在したことがある地名を時系列で記入してください。

<table>
<thead>
<tr>
<th>滞在地域</th>
<th>滞在期間</th>
</tr>
</thead>
<tbody>
<tr>
<td>都道府県名</td>
<td>市町村名</td>
</tr>
<tr>
<td>a.</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td></td>
</tr>
</tbody>
</table>

（11）今までに特別な発声訓練を受けたことがありますか。
（1）（はい・いいえ）
（2）「はい」と回答された方にお聞きします。どんな種類の発声訓練ですか。あてはまるものすべてに○をしてください。

a）演説・司会など
b）声楽
c）演技

d）演説・司会など

（12）現在、どなたかとご一緒にお住まいですか。
（1）（はい・いいえ）
（2）「はい」と回答された方にお聞きします。どなたとお住まいですか。あてはまるものすべてに○をしてください。

a）配偶者
g）祖父
b）子供
h）孫
c）母親
i）兄弟・姉妹
d）父親
j）その他の親戚など
e）嫁・婿（子供の配偶者）
k）友人
f）祖母
l）その他

（13）最終学歴（現在学生の方は現在の所属教育機関）
a）中学校
e）大学
b）高等学校
f）大学院
c）高等専門学校
g）専門学校
d）短期大学
h）その他（        ）
B. 身体活動量

(1) 仕事中のあなたの身体活動量は、あなたと同世代で同姓の人と比べると次のどれにあてはまりますか。いずれかひとつに○をしてください。
   a) かなり活動的
   b) やや活動的
   c) 相同程度
   d) やや活動的でない
   e) かなり活動的でない
   f) 該当なし

(2) 仕事以外での活動において、あなたと同世代で同姓の人と比べるとあなたの身体活動量は次のどれにあてはまりますか。いずれかひとつに○をしてください。
   a) かなり活動的
   b) やや活動的
   c) 相同程度
   d) やや活動的でない
   e) かなり活動的でない
   f) 該当なし

(3) 現在 31 歳以上の方にお聞きします。30 歳以下の方は質問 (4) に進んでください。今までのあなたの身体活動量を振り返ってみて下さい。当時の同世代で同姓の人と比べると、あなたの身体活動量は次のどれにあてはまりますか。各世代に関して、あてはまる番号をひとつ次の中から選んで下さい。
   a) かなり活動的
   b) やや活動的
   c) 相同程度
   d) やや活動的でない
   e) かなり活動的でない

   (1) 20 代 ________
   (2) 30 代 ________
   (3) 40 代 ________
   (4) 50 代 ________
   (5) 60 代 ________
   (6) 70 代 ________
(4) 定期的に激しい運動や肉体労働に従事していますか。
   (1) (はい・いいえ)
   (2) 「はい」と回答された方にお聞きします。激しい運動や肉体労働を週に
       3回以上しますか。（はい・いいえ）

(5) あなたは、どのくらいの頻度で激しい運動や活動をしますか。（例：ジョギング、エアロビクス、水泳、重い物を持ち上げたり運んだりする、畑仕事をする、
       早足で歩くこと、のぼり坂で自転車に乗る等）いずれかひとつに○をして下さい。
       a) ほぼ毎日
       b) 週に3～4回
       c) 週に1～2回
       d) （ほとんど）まったくしない

(6) あなたは、どのくらいの頻度で適度な運動や活動をしますか。（例：軽スポーツ、
       ゴルフ、社交ダンス、簡単な庭作業や大工仕事、平坦な道で自転車に乗る、
       洗車、床みがき等）いずれかひとつに○をして下さい。
       a) ほぼ毎日
       b) 週に3～4回
       c) 週に1～2回
       d) （ほとんど）まったくしない

(7) 次の項目の中で、他の方の手伝いを必要とするものがありますか。あてはま
       るもの全てに○をしてください。
       a）入浴
       b）身支度
       c）歩行
       d）摂食
       e）排泄

(8) 毎日何分くらい歩きますか。  一日につき___________分

(9) あなたは、通常どのくらいの速さで歩きますか。（いずれかに○）
       a）ゆっくり
       b）普通
       c）速い
       d）かなり速い

(10) 毎日何段くらい階段を上がりますか。
       一日につき________段（一階分を10段と考えてください。）
C. 嗜好品

(1) 現在、煙草を吸われていますか。
   (1) (はい・いいえ)
   (2) 「はい」と回答された方にお聞きします。
       a) 一日平均何本くらい喫煙しますか。 _____本
       b) 何歳から喫煙されていますか。 _____歳

(2) 喫煙をしていたことがありますか。
   (1) (はい・いいえ)
   (2) 「はい」と回答された方にお聞きします。
       a) 禁煙をしてから何年くらいですか。 _____年
       b) 禁煙するまでに、何年くらい喫煙されていましたか。_____年

(3) 以下の飲み物をどれくらいの頻度で飲まれますか。各項目につき、ひとつ○をして下さい。

<table>
<thead>
<tr>
<th>頻度</th>
<th>ほと んど 飲ま ない</th>
<th>月に 1〜2 杯</th>
<th>週に 1〜2 杯</th>
<th>週に 3〜6 杯</th>
<th>1 日 1〜2 杯</th>
<th>1 日 3〜5 杯</th>
<th>1 日 6 杯 以上</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. カフェイン入りのソ フトドリンク（コーラ等）</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. 珈琲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. お茶（日本茶・紅茶等）</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. ビール等</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. ワインや日本酒等</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. ウイスキーやブラン デー</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
D. 健康状態

ここでは、個人的な質問をさせていただきます。ここでお伺いする質問内容は、学術研究のためだけに使用されるものであり、ご記入内容が第三者に漏れるところはありません。なるべく全ての質問にお答え頂きたいと思いますが、お答えにくい部分は無回答でかいません。何卒、ご協力をお願いいたします。

| (1) 発話障害があると診断されたことが在りますか。 (はい・いいえ) |
| (2) 言語障害があると診断されたことが在りますか。 (はい・いいえ) |
| (3) 普段、思うように発音できないと感じることがありますか。 (はい・いいえ) |
| (4) 他の人からあなたが言ったことを聞き返されることがよくありますか。 (はい・いいえ) |
| (5) お仕事や娯楽活動において、大きな声をずっと出し続けたことがありますか。 (はい・いいえ) |
| (6) 最近、お仕事や娯楽活動において、大きな声をずっと出し続けたことがありますか。 (はい・いいえ) |
| (7) ご家族に発話・言語障害をお持ちの方がいらっしゃいますか。 (はい・いいえ) |
| (8) 聴覚障害があると診断されたことが在りますか。 (はい・いいえ) |
| (9) 御自分の聴覚に問題があるとお考えですか。 (はい・いいえ) |
| (10) 慢性的な耳の痛みがありますか。 (はい・いいえ) |
| (11) ご家族に、難聴などの聴覚障害の方はいらっしゃいますか。 (はい・いいえ) |
| (12) お仕事や娯楽活動などにおいて、常に大きな音や騒音にさらされていましたことがありますか。 (はい・いいえ) |
| (13) めまいがしたり、耳鳴りがしたりしますか。 (はい・いいえ) |
| (14) あなたは、若い頃、言語療法を受けたことがありますか。 (はい・いいえ) |
| (15) 欠損歯はありますか。 |
| (1) (はい・いいえ) |
| (2) 「はい」と回答された場合、欠損歯の数を記入してください。 |
| a) 上前歯  ________本 |
| b) 下前歯  ________本 |
| c) 前歯以外 ________本 |
| (16) 架橋義歯はありますか。 (はい・いいえ) |
| (17) 義歯（入れ歯）はありますか。 |
| (1) (はい・いいえ) |
| (2) 「はい」と回答された場合、義歯をしている箇所はどこですか。あてはまるものを〇をしてください。 |
| a) 上歯 |
| b) 下歯 |
c) 上下とも
(18) 規則的な歯並びをしていますか。（はい・いいえ）

(19) あなたの健康状態は？（一番よくあてはまるものに○印をつけて下さい）
    a) よくない
    b) あまりよくない
    c) 普通
    d) よい
    e) とてもよい

(20) 医師から以下の疾患・症状などがあると診断されたことがありますか。ある場合は、発症年齢を記入してください。
    a) 冠状動脈性心臓病 （はい・いいえ）____歳
    b) 心臓不整脈 （はい・いいえ）____歳
    c) 脳卒中 （はい・いいえ）____歳
    d) 高血圧 （はい・いいえ）____歳
    e) 慢性気管支炎 （はい・いいえ）____歳
    f) ぜんそく （はい・いいえ）____歳
    g) 慢性鼻炎 （はい・いいえ）____歳
    h) 慢性的な耳の疾患 （はい・いいえ）____歳
    i) うつ病 （はい・いいえ）____歳
    j) その他の神経性疾患 （はい・いいえ）____歳

(21) のどを損傷あるいは手術を受けたことがありますか。（はい・いいえ）
(22) 以下の薬を常用していますか。（あてはまるもの全てに○）
    a) 抗ヒスタミン剤（アレグラ・クラリチン・パブロン鼻炎薬など）
    b) アンフェタミン（アデロールなど）
    c) 鎮静薬
    d) 鎮痙薬として作用する薬剤（アトロピン・スコポラミン・ジフェノキシレート系の薬剤）
    e) 病院から薬をもらって飲んでいるが、薬の名前は分からない。

男性の方への質問は以上です。
女性の方は以下の質問（23）にお答えください。

(23) 月経はありますか。
    (1) （ない ある）
    (2) 「ある」と答えた方にお聞きします。月経の周期は規則的ですか。
        a) 規則的
        b) 不順
Appendix C1: Recording Instructions (English version). All text on the instruction sheets used in the recording was written in an 18-point font size for better visibility.

General Instructions for Recording

If you are having any difficulty reading this sentence, please let me know now.

You will be asked to pronounce vowels by themselves, and then to read several sentences and two passages. You will be given specific instructions later. To obtain a good quality of recording, I would like to ask you the following things before I begin your recording.

Before the recording:

- Please turn off your cellular phone or pager if you have one.
- Please turn off any alarms.
- Please use a rest room before the recording.

During the recording:

- Please try not to touch the microphone. Please tell me when you feel uncomfortable.
- You can drink water whenever you wish.

Do you have any questions?
Instructions for sustained vowels 1

I will ask you to pronounce vowels “ah” as in “ba-ba black sheep” and “ee” as in “bee keeper” several times.

Please pronounce each vowel in your comfortable speaking voice, and sustain your pronunciation until I tell you to stop.

Please try to pronounce the vowel as steadily and clearly as possible. Keep the same pitch and loudness levels throughout each pronunciation.

We will start with the vowel “ah”, and then do the same thing for “ee”. Take a deep breath before you begin each pronunciation. Now let’s practice your pronunciation before the recording. When I point to you, please start saying the vowel “ah”. Please tell me when you are ready.

Remember, take a deep breath each time you start pronouncing the vowel. You will be asked to pronounce each vowel at your comfortable pitch and loudness level for three times.

Do you have any questions?
Instructions for sustained vowels 2

This time I want you to pronounce the same vowels in just a slightly different way. When you speak, you make some low pitch sounds and some high pitch sounds. You are to pronounce the same vowels “ah” and “ee” at a high pitch level that is still comfortable, and sustain your pronunciation until I tell you to stop.

Please try to pronounce the vowel as steadily and clearly as possible. Keep the same pitch and loudness level during each pronunciation. We will start with the vowel “ah” and then do “ee”. Now let’s practice your pronunciation before the recording. When I point to you, please start saying the vowel “ah”. Tell me when you are ready.

Remember, take a deep breath each time you start pronouncing the vowel. You will be asked to pronounce each vowel for three times.

Do you have any questions?
Instructions for sustained vowels 3

This time, I want you to pronounce the same vowels at a low pitch level that is still comfortable, and sustain your pronunciation until I tell you to stop.

Again, try to pronounce the vowel as steadily and clearly as possible, keeping the same pitch and loudness level during each pronunciation. We will start with the vowel “ah” and do “ee”. Now let’s practice your pronunciation before the recording. When I point to you, please start pronouncing the vowel “ah”. Tell me when you are ready.

Remember, take a deep breath each time. We will do this three times for each vowel.
Instructions for reading sentences 1

This time, I have several short sentences written on cards. You see the sentences below. Please take a moment to read through the list, especially reading the letters.

You will see one of the sentences on a card that I’m going to show you. When I show you a card, please start to read the sentence aloud. Say all the sentences in the same way. If you stumble or don’t like your pronunciation, just repeat the whole sentence from the beginning as many times as you like. Please don’t read the numbers. Do you have any questions?

Sentence list:

1. The answer was the ABC Company.
2. The answer was the ACC Company.
3. The answer was the BBC Company.
4. The answer was the BCC Company.
5. The answer was ABC Corporation.
6. The answer was ACC Corporation.
7. The answer was BCC Corporation.
8. The answer was BBC Corporation.
Instructions for reading sentences 2

This time, you are going to read a short dialogue between two people. Imagine the following situation:

Two people (Pat and Terry) are watching a popular quiz show on TV. Today’s show has various questions about the famous companies. In the last question, the respondents were asked, “Which one of the companies remains atop the industry?” One contestant gave an answer. But, Pat didn’t catch the answer for this question. Pat will ask Terry about the answer. Terry will tell Pat the answer.

You can see examples of a dialogue below.

Example 1

Pat: What was the answer?
Terry: The answer was BCC Corporation.

Example 2

Pat: Was it BBC?
Terry: No, the answer was BCC Corporation.

You will see a short dialogue like these examples on a card that I’m going to show you. When I show you a card, start to read the sentences aloud. Please don’t read the names. If you stumble or don’t like your pronunciation, please repeat the dialogue from the beginning as many times as you like.

Do you have any questions?
Instructions for reading sentences 3

You will be asked to read a sentence written on a card. This time, I would like you to read the sentence as if you are younger or older than you are. The card will say either “20 years younger” or “20 years older”.

When I point to you, please start to read the sentence aloud. If you stumble or don’t like your pronunciation, just repeat the sentence from the beginning as often as you need. Do you have any questions?

Sample cards:

[20 years Younger]
The answer was BCC Corporation.

[20 years Older]
The answer was BCC Corporation.
Instructions for reading sentences 4

You will be asked to read the short sentences below about five bears (Baby Bear, Mama Bear, Papa Bear, Grandma Bear, and Grandpa Bear). I would like you to read the sentence pretending to be the speaker. For example, when you read the sentence by Baby Bear, I would like you to read the sentence pretending you are Baby Bear. Or, when you read the sentence by Papa Bear, you should pretend to be Papa Bear. You are going to imitate all five different voices.

When I point to you, please start to read the sentences. If you stumble or don’t like your pronunciation, please repeat the sentence from the beginning as much as you like.

Do you have any questions?

When Baby Bear talks, he sounds like this,
“The answer was BCC Corporation.”

When Mama Bear talks, she sounds like this,
“The answer was BCC Corporation.”

When Papa Bear talks, he sounds like this,
“The answer was BCC Corporation.”

When Grandma Bear talks, she sounds like this,
“The answer was BCC Corporation.”

When Grandpa Bear talks, he sounds like this,
“The answer was BCC Corporation.”
Instructions for reading a passage

Now, you will be given two passages of material to read. I want you to read each passage two times. After you have completed reading the passage for the first time, take a short break, and go back to read it aloud for a second time. This time, if you miss a word, please don’t stop but continue reading to the end of the passage.

Please take a moment to carefully read through the first passage in silence. Do you have any difficulty reading the passage or are there any words you cannot pronounce?

Do you have any questions?

When I point to you, please begin reading the passage from the title of the passage. Remember, don’t worry if you miss a word, just keep reading on through the passage.
The North Wind and the Sun

The North Wind and the Sun were arguing one day about which of them was stronger, when a traveler came along wrapped up in an overcoat. They agreed that the one who could make the traveler take his coat off would be considered stronger than the other one.

Then the North Wind blew as hard as he could, but the harder he blew, the tighter the traveler wrapped his coat around him; and at last the North Wind gave up trying. Then the Sun began to shine hot, and right away the traveler took his coat off. And so the North Wind had to admit that the Sun was stronger than he was.

(After a short break, please read again from the title.)
The Rainbow Passage

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow. Throughout the centuries people have explained the rainbow in various ways. Some have accepted it as a miracle without physical explanation. To the Hebrews it was a token that there would be no more universal floods. The Greeks used to imagine that it was a sign from the gods to foretell war or heavy rain. The Norsemen considered the rainbow as a bridge over which the gods passed from earth to their home in the sky. Others have tried to explain the phenomenon physically. Aristotle thought that the rainbow was caused by reflection of the sun's rays by the rain. Since then physicists have found that it is not reflection, but refraction by the raindrops which causes the rainbows. Many complicated ideas about the rainbow have been formed. The difference in the rainbow depends considerably upon the size of the drops, and the width of the colored band increases as the size of the drops increases. The actual primary rainbow observed is said to be the effect of superposition of a number of bows. If the red of the second bow falls upon the green of the first, the result is to give a bow with an abnormally wide yellow band, since red and green light when mixed form yellow. This is a very common type of bow, one showing mainly red and yellow, with little or no green or blue.

(After a short break, please read again from the title.)
録音実験の概要（事前説明）

この文が読みづらい場合は、担当者にお伝え下さい。

これから、みなさんに母音を発音したり、いくつかの文や文章を読んだりして頂きます。それぞれの課題は、のちほど具体的に説明させて頂きます。良質な録音を得るために、録音を始める前にみなさんに次の事項に関してお願いがあります。

実験の前に

- 携帯電話をお持ちの方は、電源をお切り下さい。
- 腕時計などのアラーム機能をお切り下さい。
- トイレに行きたい方は、今のうちに行っておいて下さい。

実験中

- 実験開始後は、ヘッドセットマイクには手を触れないようにして下さい。
- マイクの付け心地が悪い場合は、担当者にお知らせ下さい。
- お水を飲みたい場合は、いつでも飲むことができます。
- 休憩をとりたい時は、手を上げてお知らせ下さい。

何かご質問はございますか。
課題 1（母音の録音 1）の説明

これから母音の「あ」と「い」を何度か言って頂きます。それぞれの母音をいつもの会話で使っているように発音して、そのまま私が合図するまで発音し続けて下さい。

母音はできるだけはっきりと同じ調子で発音して下さい。発音している間は、声の高さや大きさは変化させないようにして下さい。

まず、母音の「あ」からはじめ、その後同様のことを「い」でも行います。母音を発する前には、必ず深く息を吸い込むようにして下さい。実際の録音をする前に練習をしてみましょう。私が合図したら、「あー」と言い始めて下さい。準備はいいですか。

それでは、これからそれぞれの母音を3回ずつ言って頂きます。母音を言い始める前には必ず深く息を吸い込むようにして下さい。

何かご質問はございますか。
課題2（母音の録音2）の説明

今度は、先ほどと同じ母音「あ」と「い」を少し言い方を変えて発音して頂きます。誰でも会話をする時には、高い声を使ったり、低い声を使ったりします。今回は、それぞれの母音を、あなたが無理をしない範囲の高い声で発音して下さい。先ほどと同じように、それぞれの母音は、私が合図するまで発音し続けて下さい。

母音はできるだけはっきりと同じ調子で発音して下さい。発音している間は、声の高さや大きさは変化させないようにして下さい。

まず、母音の「あ」からはじめ、その後同様のことを「い」でも行います。母音を発する前には、必ず深く息を吸い込むようにして下さい。実際の録音をする前に練習をしてみましょう。私が合図したら、「あー」と言い始めても良いです。

それでは、これからそれぞれの母音を3回ずつ言って頂きます。母音を言い始める前には必ず深く息を吸い込むようにして下さい。

何かご質問はございますか。
課題3（母音の録音3）の説明

今度は、同じ母音「あ」と「い」を低い声で発音して頂きます。それぞれの母音を、あなたが無理をしない範囲の低い声で発音して下さい。先ほどと同じように、それぞれの母音は、私が合図するまで発音し続けて下さい。

母音はできるだけはっきりと同じ調子で発音し、声の高さや大きさは変化させないようにして下さい。

まず、母音の「あ」からはじめ、その後同様のことを「い」でも行います。母音を発する前には、必ず深く息を吸い込むようにして下さい。実際の録音をする前に練習をしてみましょう。私が合図したら、「あー」と言い始めます。準備はいいですか。

それでは、これからそれぞれの母音を3回ずつ言って頂きます。母音を言い始める前には必ず深く息を吸い込むようにして下さい。
課題4（文の録音1）の説明

今度は、カードに書かれたいくつかの短文を読んで頂きます。読んで頂く文は、以下に列挙してあります。特に英字の部分に注意して、文の一覧に目を通して下さい。

それぞれの文は、別々のカードに書かれています。私がカードをお見せしたら、カードに書かれてある文を声を出して読んで下さい。全ての文は、同じように読むようにして下さい。もし、言い間違えたり、ご自分の発話が気にいらないう場合は、何度でも言い直してかまいません。言い直したい場合は、文の最初から言い直して下さい。文の前の番号は読む必要はありません。

何かご質問はございますか。

文の一覧
1. 答えは、エービーシー・カンパニーだったよ。
2. 答えは、エーシーシー・カンパニーだったよ。
3. 答えは、ビービーシー・カンパニーだったよ。
4. 答えは、ビーシーシー・カンパニーだったよ。
5. 答えは、エービーシー・コーポレーションだったよ。
6. 答えは、エーシーシー・コーポレーションだったよ。
7. 答えは、ビービーシー・コーポレーションだったよ。
8. 答えは、ビーシーシー・コーポレーションだったよ。
課題 5（会話文の録音）の説明

今度は、山田さんと鈴木さんの短い会話文を読んで頂きます。次に述べるような状況を想像して下さい。

『山田さんと鈴木さんは、人気のクイズ番組をテレビで見ています。今日の番組は、有名な会社に関した質問をしています。前の問題は、「次に挙げる会社の中で、どの会社が業界第一に位置するでしょう？」でした。回答者の一人が答えを当てましたが、山田さんは聞き逃してしまいました。山田さんは、前の問題の答えを鈴木さんに聞いています。すると、鈴木さんは、山田さんに答えを教えてあげます。』

会話文の例を以下に示します。

会話文例1
山田：答えは何だった？
鈴木：答えは、エービーシー・コーポレーションだったよ。

会話文例2
山田：答えは、エービーシー・コーポレーションだった？
鈴木：いや、答えは、ビービーシー・コーポレーションだったよ。

このような会話文が、これからお見せするカードに書かれています。私がカードをお見せしたら、カードに書かれてある文を声を出して読んで下さい。名前の部分は読まないで下さい。もし、言い間違えたり、ご自分の発話が気にいる場合は、何度でも言い直してかまいません。言い直したい場合は、文の最初から言い直して下さい。
何かご質問はございますか。

何かご質問はございますか。
課題 6（文の録音 3）の説明

今度もカードに書かれた文を読んで頂きます。今回は、文を読む時に、ご自分が 20 歳若くなったつもり、あるいは、20 歳年を取ったつもりで文を読んで下さい。カードに、「20 歳若く」あるいは「20 歳年を取って」と指示してあります。

私がカードをお見せしたら、カード上の指示に従って、文を読んで下さい。もし、言い間違えたり、ご自分の発話が気にいらない場合は、何度でも言い直してかまいません。言い直したい場合は、文の最初から言い直して下さい。

何かご質問はございますか。

カードの例:

[20 歳若く]

答えは、ビーシーシー・コーポレーションだったよ。

[20 歳年を取って]

答えは、ビーシーシー・コーポレーションだったよ。
課題 7（文の録音 4）の説明

今回は、以下に示す 5 匹の熊（赤ちゃん熊・お母さん熊・お父さん熊・おばあさん熊・おじいさん熊）の文を読んで頂きます。今回は、それぞれの話し手のなったつもりで文を読んで下さい。例えば、赤ちゃん熊が話している文は、赤ちゃん熊になったつもりで読んで下さい。同様に、お父さん熊が話している文は、お父さん熊になったつもりで読んで下さい。

私がカードをお見せしたら、文を読み始めて下さい。もし、言い間違えたり、ご自分の発話が気にいらない場合は、何度でも言い直してかまいません。言い直したい場合は、文の最初から言い直して下さい。

何かご質問はございますか。

赤ちゃんグマがしゃべると、こんな風に聞こえます。
「答えは、ビーシーシー・コーポレーションだったよ。」

お母さんグマがしゃべると、こんな風に聞こえます。
「答えは、ビーシーシー・コーポレーションだったよ。」

お父さんグマがしゃべると、こんな風に聞こえます。
「答えは、ビーシーシー・コーポレーションだったよ。」

おばあさんグマがしゃべると、こんな風に聞こえます。
「答えは、ビーシーシー・コーポレーションだったよ。」

おじいさんグマがしゃべると、こんな風に聞こえます。
「答えは、ビーシーシー・コーポレーションだったよ。」
課題8（文章の録音）の説明

今から、2種類の文章が書かれた用紙をお渡します。今回は、それぞれの文章を2回ずつ読んで頂きます。文章を読み終えたら、少し休憩してから同じ文章をもう一度最初から読んで下さい。今回は、言い間違えたり、読み間違えたりしても、止まらずにそのまま最後まで読み続けてください。

まず、最初の文章を声を出さずに丁寧に読んでみてください。読みづらい文章や、読み方を確認したい単語があれば担当者にお尋ね下さい。

私が合図したら、題名から読みはじめて下さい。仮に、読み間違えても気にしないで、最後まで読み続けて下さい。

何かご質問はございますか。
きたかぜ

北 風 と太陽

ある時、北 風 と太陽が 力 くらべをしました。旅 人の 外 套 を脱がせた方が勝ちということに決めて、まず 北 風 から始めました。北 風 は、「なに、一 ひとまくりにして見せよう。」と、激しく吹き立てました。すると 旅 人は、

北 風 が吹けば吹くほど 外 套 をしっかりと 体 にくっつけました。今度は太陽の番になりました。太陽は雲のあいだから優しい顔を出して暖かな光を送りました。旅 人は段々よい 心 もちになって、しまいには 外 套 を脱ぎました。

そこで 北 風 の負けになりました。

(少し休憩して、題名の部分からもう一度読んで下さい。)
虹の話

太陽の光が空中の雨粒にあたると、雨粒がプリズムのように働き、虹を形成する。虹は、白色光が多くの美しい色に分散しておこる。これらの分散した色は、上空に大きな半円形として現われ、その両端は地平線の上に位置する。伝説によれば、虹のかたたには黄金が湧いて出る壺があるといつも。この壺を探した者はいたが、見つけた者は誰もいない。だから誰かが無謀に何かを求めていると、その友人達は、あいつは虹の向こうの黄金の壺を探しているんだ、と言ったりする。

何世紀もの間、虹に関して様々な説明が出されてきた。ある者は物理的な説明を求めず、単に奇跡としてそれを受け入れた。ヘブライ人にとって、虹はこれ以上大きな洪水がないという予告であった。ギリシャ人は虹を戦争か大雨を予告する神からの神託と考えた。

ノルウェーの人々は、虹を神々が地上から天空の住まいへと戻るための架け橋と考えた。またある者は、この現象を物理的に説明することを試みた。アリストテレスは、虹は太陽光線が雨に反射することによって引き起こされると考えた。しかしそれ以降、物理学者達は、虹は反射ではなく、雨による光の屈折である
ことを見た。虹について、多くの複雑なしくみがわかってきた。虹の
見え方は、雨粒の大きさに非常に左右され、虹の幅は、雨粒が大きくなるに
つれて広くなる。実際に観察される主虹は多くの光の帯が重ね合わされた結果で
あると言われる。二番目の赤色の帯に、一番目の緑色の帯が重なれば、赤色
と緑色が混合して黄色になるため、結果として異常に幅広い黄色い光の帯が生じる。こういった主に赤色と黄色からなり、緑色と青色がほとんど見えない虹は非常によく見られる。
Appendix D1: Listener Questionnaire (English version). The actual response sheets were written in a 14-point font size.

Please answer the following questions. Please ask me if you would like any assistance in filling the questionnaire out.

A. Background Information

(1) Sex: Male ____ Female ____

(2) Ethnicity:
   a) Hispanic or Latino__
   b) Not Hispanic or Latino __

(3) Race:
   a) American Indian or Alaska Native __
   b) Asian __
   c) Black or African American __
   d) Native Hawaiian or other Pacific Islander __
   e) White __

(4) Date of birth: Month___ Day___Year 19____ Age____ years old

(5) What is your occupation/Major?
   (1) ______________________
   (2) If retired, what was your occupation? ______________________________

(6) What is your native language? ______________________________

(7) Do you know any other language(s)?
   (1) Yes____ No____
   (2) If yes, please list the language(s) you know.

(8) Do you use any other language(s) on a regular basis?
   (1) Yes____ No____
   (2) If yes, please list the language(s) you use. ______________________________

(9) Where do you come from?
City_________________________ State____________

(10) Please list the place(s) you have lived more than 6 months in chronological order.

<table>
<thead>
<tr>
<th>City name, State</th>
<th>Duration (from Year to Year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td></td>
</tr>
<tr>
<td>b)</td>
<td></td>
</tr>
<tr>
<td>c)</td>
<td></td>
</tr>
<tr>
<td>d)</td>
<td></td>
</tr>
<tr>
<td>e)</td>
<td></td>
</tr>
</tbody>
</table>
(11) Do you live with somebody else?
(1) Yes____ No____
(2) If yes, whom do you live with? (Please check all that apply.)
   (a) Spouse ___
   (b) Child(ren) ___
   (c) Mother ___
   (d) Father ___
   (e) Child(ren)’s spouse ___
   (f) Grandfather ___
   (g) Grandmother ___
   (h) Grandchild(ren) ___
   (i) Brother(s)/Sister(s) ___
   (j) Other family members ___
   (k) Friend(s) ___
   (l) Other ___

(12) What is your highest education level completed (or currently working on)?
   (a) Less than high school ___
   (b) High school Diploma ___
   (c) Associates Degree ___
   (d) Bachelor’s Degree ___
   (e) Masters Degree ___
   (f) Doctoral Degree ___
   (g) Professional Degree ___
   (h) Other: __________________
B. Social Habits

(1) Do you smoke cigarettes?
   (1) Yes ____ No ____
   (2) If Yes, how many on an average day? ________
   (3) Age started? ________

(2) Did you ever smoke cigarettes?
   (1) Yes ____ No ____
   (2) If Yes, How many years have passed since you stopped smoking? _____
   (3) How many years did you smoke before you stopped smoking? ______

C. Past and Present Health Status

In this section of the survey, we'd like to ask you about your personal information. Remember, we do not want you to answer any questions if you are uncomfortable doing so. Information you will provide for the current research is collected and used only for academic purposes, and will not be shared by anybody other than the principle investigator.

(1) Have you ever been diagnosed with a speech problem?  Yes ____ No ____
(2) Have you ever been diagnosed with a language problem? Yes ____ No ____
(3) Have you noticed difficulty with your speech? Yes ____ No ____
(4) Have you been often asked to repeat what you said? Yes ____ No ____
(5) Is there a family history of speech/language problem? Yes ____ No ____
(6) Have you ever been diagnosed with a hearing problem? Yes ____ No ____
(7) Have you noticed difficulty with your hearing? Yes ____ No ____
(8) Do you have a history of ear infection or any pain in your ears? Yes ____ No ____
(9) Is there a family history of hearing loss? Yes ____ No ____
(10) Do you have any history of exposure to noise in recreational activities, at work, or in the military? Yes ____ No ____
(11) Do you experience dizziness or ringing (tinnitus) in ears? Yes ____ No ____
(12) Did you have any speech therapy when you were younger? Yes ____ No ____
(13) Has a physician ever told you that you had any of the following? (Please check and give year of onset, if applicable.)
   a) Stroke  Yes ____ No ____ Year ________
   b) Chronic ear diseases Yes ____ No ____ Year ________
   c) Other neurological disease(s) Yes ____ No ____ Year ________

221
Appendix D2: Listener Questionnaire (Japanese version).

以下の質問事項に回答して下さい。必要な場合は、実験担当者が記入をお手伝いしますので、お知らせ下さい。

A. 基本事項

(1) 性別：男性・女性
(2) 民族：
   a) ヒスパニック・ラテン系 b) ヒスパニック・ラテン系以外
(3) 人種：
   a) アメリカ・インディアン及びアラスカ原住民 b) アジア系
c) 黒人・アフリカ系
   d) ハワイ原住民または太平洋諸島系
e) 白人
(4) 生年月日：大正・昭和 ______年____月____日（）歳
(5) 職業：
   (1) 現在の御職業________________________________________
   (2) 引退されている場合、以前の御職業________________________
(6) あなたの母国語は何ですか。___________________________
(7) 日本語以外の外国語を流暢に話せますか。
   (1) (はい・いいえ)
   (2) 「はい」と回答された方、よく知っている外国語をすべて書いて下さい________________________________
(8) 日本語以外の外国語を日常使用していますか。
   (1) (はい・いいえ)
   (2) 「はい」と回答された方、日常使用している外国語をすべて書いて下さい________________________________
(9) あなたの方言は何ですか。
   (1) 方言（大阪弁・京都弁など）______________________________
   (2) あなたの方言を習得した地域は主にどこですか。
       __________________府県____________________________市町村

222
（10）日本語以外の外国語が話される環境に、半年以上滞在したことがありますか。

（1）（はい・いいえ）
（2）「はい」で回答された方にお聞きします。以前に半年以上滞在したことがある地域名を時系列で記入してください。

<table>
<thead>
<tr>
<th>滞在地域</th>
<th>滞在期間</th>
</tr>
</thead>
<tbody>
<tr>
<td>国名</td>
<td>何歳から</td>
</tr>
<tr>
<td>a.</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td></td>
</tr>
</tbody>
</table>

（11）現在、どなたかとご一緒にお住まいですか。
（1）（はい・いいえ）
（2）「はい」と回答された方にお聞きします。どなたとお住まいですか。すべてはまるものすべてに○をしてください。
a）配偶者  b）子供  c）母親  d）父親  e）嫁・婿（子供の配偶者）  f）祖母  g）祖父  h）孫  i）兄弟・姉妹  j）その他の親戚など  k）友人  l）その他

（12）最終学歴（現在学生の方は現在の所属教育機関）
a）中学校  b）高等学校  c）高等専門学校  d）短期大学  e）大学  f）大学院  g）専門学校  h）その他（   ）
B. 嗜好品

(1) 現在、煙草を吸われていますか。
   (1) (はい・いいえ)
   (2) 「はい」と回答された方にお聞きします。
      a) 一日平均何本くらい喫煙しますか。 _____本
      b) 何歳から喫煙されていますか。 _____歳

(2) 喫煙をしていたことがありますか。
   (1) (はい・いいえ)
   (2) 「はい」と回答された方にお聞きします。
      a) 禁煙をしてから何年くらいですか。 _____年
      b) 禁煙するまでに、何年くらい喫煙されていましたか。_____年

C. 健康状態

ここでは、個人的な質問をさせていただきます。ここでお伺いする質問内容は、学術研究のためだけに使用されるものであり、ご記入内容が第三者に漏れることはありません。なるべく全ての質問にお答え頂きたいため、お答えにくい部分は無回答でかまいません。何卒、ご協力をお願いいたします。

(1) 発話障害があると診断されたことが在りますか。 (はい・いいえ)
(2) 言語障害があると診断されたことが在りますか。 (はい・いいえ)
(3) 普段、思うように発音できないと感じることがありますか。
    (はい・いいえ)
(4) 他の人からあなたが言ったことを聞き返されることがよくありますか。
    (はい・いいえ)
(5) ご家族に発話・言語障害をお持ちの方がいらっしゃいますか。
    (はい・いいえ)
(6) 聴覚障害があると診断されたことが在りますか。 (はい・いいえ)
(7) 御自分の聴覚に問題があるとお考えですか。
    (はい・いいえ)
(8) 慢性的な耳の痛みがありますか。
    (はい・いいえ)
(9) ご家族に、難聴などの聴覚障害の方はいらっしゃいますか。
    (はい・いいえ)
(10) お仕事や娯楽活動などにおいて、常に大きな音や騒音にさらされてきたことがありますか。
    (はい・いいえ)
(11) めまいがしたり、耳鳴りがしたりしますか。
    (はい・いいえ)
(12) あなたは、若い頃、言語療法を受けたことがありますか。
    (はい・いいえ)
医師から以下の疾患・症状などがあると診断されたことがありますか。ある場合は、発症年齢を記入してください。

a）脳卒中 （はい・いいえ）______歳

b）慢性的な耳の疾患 （はい・いいえ）______歳

c）その他の神経性疾患 （はい・いいえ）______歳
Curriculum Vitae

Education

Ph.D. Indiana University, Bloomington, IN, USA, September, 2006
(First major: Linguistics)
(Second major: Speech and Hearing Sciences)

Dissertation: Cross-language study of age perception
Advisors: Kenneth de Jong and Diane Kewley-Port

M.A. Indiana University, Bloomington, IN, USA, 1999
(Major: Linguistics)

M.A. Konan University, Kobe, Japan, 1997
(Major: English (Linguistics))

Stanford University, Stanford, CA, USA, Summer, 1994
A summer program on Sociology and Sociolinguistics

B.A. Kwansei Gakuin University, Nishinomiya, Japan, 1995
(Major: Literature)

Awards and Fellowships

- Indiana University Doctoral Student Grant-in-Aid of Research Award (2005)
- IU Speech & Hearing Sciences Departmental Research Support Award (2005)
- IU Graduate & Professional Student Organization Research Award (Fall, 2004)
- Fred W. Householder Memorial Awards (2001, 2005)
- International and Interdisciplinary Research Conference on Aging and Speech Communication student scholarship (2005)
- Indiana University Women in Science Program Travel Grant (2006)
- Indiana University Linguistic Club Travel Grants
- Konan University Fellowship, Konan University
- Kwansei Gakuin University Fellowship, Kwansei Gakuin University

Certification


Refereed Papers


Working Papers


**Presentations**


Nagao, K. (to be presented). Cross-language study of age stereotypes in speech perception. To be presented at the NWAV35 on November 10 at Columbus, OH.


**Research Experience**

**Research assistant**, Indiana University, 2004-Present.
- Data process for the on-going project at the Linguistic Speech Laboratory
- Project: Prosody in cross-language perception and production
- Supervisor: Kenneth de Jong

**Graduate research assistant**, Indiana University, 2000 – 2003.
- Design and conduct experiments at the Linguistic Speech Laboratory
- Project: Production and perception effects on syllabic affiliation
- Supervisor: Kenneth de Jong

• Design and conduct experiments
• Project: Gender differences on the perception of emphasized words
• Supervisor: Shigeaki Amano

• Data analysis of epenthetic vowels in speech by Japanese learners of English
• Primary investigator: Keiichi Tajima

Graduate research assistant, Konan University, 1995 – 1997.
• Data analysis of lip movements for Japanese vowel
• Director: Yoshihiro Masuya

Teaching Experience
Assistant instructor, Indiana University, Introductory of Phonetics L541, Spring 2004. Taught discussion sections, supervised lab projects.
Volunteer teaching of Japanese to beginner and intermediate students

Teaching English to Japanese elementary students

Electronic and Textual Editing Experience
Associate editor, IULC Working Papers Online, 2000 – 2002
• Review and solicit manuscripts for possible publication
• Process manuscripts submitted for publication

Professional Societies

Language Competencies
Japanese (Native)       English (Advanced)
French (Reading)        Chinese (Elementary)

Computer Skills
Microsoft Word, Excel, PowerPoint, Macromedia, SPSS
Programming Languages: Matlab, AWK, C, Prologue, HTML

Professional and University Services
Webmaster, Indiana University Linguistics Speech Laboratory, 2001-2004
• Launch and maintenance of a website for the laboratory directed by Kenneth de Jong
• URL: http://www.indiana.edu/~lsl/home.html

Volunteer speaker of the Global Speakers Services, 1997-2004
• Talks and demonstrations on topics in Japanese language and culture at elementary schools, high schools, and retirement houses in central Indiana.