# Age estimation in foreign-accented speech by native and non-native speakers


## Abstract

Current research shows that listeners are generally accurate at estimating speakers' age from their speech. This study investigates the effect of speaker first language and the role played by such speaker characteristics as fundamental frequency and speech rate. In this study English and Japanese first language speakers listened to English- and Japanese accented English speech and estimated the speakers' age. We find the highest correlation between real and estimated speaker age for English listeners listening to English speakers, followed by Japanese listeners listening to both English and Japanese speakers, with English listeners listening to Japanese speakers coming last. We find that Japanese speakers are estimated to be younger than the English speakers by English listeners, and both groups of listeners estimate male speakers and speakers with a lower mean fundamental frequency to be older. These results suggest that listeners rely on sociolinguistic information in their speaker age estimations and language familiarity plays a role in their success.



## Keywords

Age estimation, foreign-accented speech, perception, sociolinguistic variation

# 1. Introduction

Listeners are generally fairly accurate in estimating speakers' age from their speech (Krauss, Freyberg, & Morsella, 2002; Moyse, 2014; Ryan & Capadano, 1978; however, see Benjamin, 1992, for a discussion to the contrary). Ryan & Capadano (1978) recorded correlations of 0.90 and 0.93 between estimated age and real age for a passage-length stimulus and Krauss et al. (2002) - a correlation of 0.61 for sentence-length stimuli. This holds true for very short extracts or degraded signal as well: listeners exhibited a 78% success rate in sorting prolonged vowels into two broad age categories (Ptacek & Sander, 1966) and could assign a speaker recorded over a telephone line to a decade-range age group with a correlation of 0.77 (Cerrato, Falcone, & Paolini, 2000).

Such robustness may be due to listeners' reliance on their knowledge of universal relationships between speaker age and acoustic correlates such as fundamental frequency (F0), rate of speech, and voice tremor (Decoster & Debruyne, 1996; Harnsberger, Brown, Shrivastav, & Rothman, 2010; Harnsberger, Shrivastav, Brown, Rothman, & Hollien, 2008; Hartman, 1979; Linville, 2001; Ramig, 1983, 1986; Ramig & Ringel, 1983; Ryan & Burk, 1974; Shipp, Qi, Huntley, & Hollien, 1992; Waller, Eriksson, & Sörqvist, 2015). For example, a 20-25% decrease in speech rate is reported for older speakers, with a smaller change for females (Schötz, 2007). The relationship between speaker age and acoustic measures is often complex. Some studies, for example, have shown that F0 decreases with speaker age while others demonstrate that it is non-linear and stays stable with a drop at menopause for women and an increase in older age for men (see Schötz, 2007, for a discussion of age-related changes and a number of acoustic correlates). If listeners solely rely on acoustic characteristics determined by physiological changes due to aging in estimation of speaker age, then we would expect them to be equally good at estimating

the age of speakers from different social backgrounds. If sociolinguistic information (which goes beyond physiological factors) is additionally involved, then we can expect to see differences in accuracy across different social groups.

Linguistic studies have demonstrated a connection between speaker age and language-specific sociolinguistic variables in both production and perception (e.g., Drager, 2010; Walker, 2007). For example, Drager (2010) found that assumed age of speaker affected vowel categorization in vowels involved in a chain shift. Moreover, speakers are perceived to be younger when producing innovative variants compared to more conservative variants (Walker, 2007). This suggests that listeners may be relying on both physiologically-determined and sociophonetic acoustic correlates of speaker age.

A number of listener biases in speaker age estimation have been attested. Several studies have reported general tendencies for underestimation (Hartman, 1979; Hughes & Rhodes, 2010; Mulac & Gilles, 1996) and regression to the middle age (Braun & Cerrato, 1999; Cerrato et al., 2000; Hunter, Ferguson, & Newman, 2016). It has also been demonstrated that listener and speaker social characteristics such as age may come into play (Huntley, Hollien, & Shipp, 1987). For example, Moyse, Beaufort, and Brédart (2014) have shown an "own-age bias" for older listeners such that they were more accurate with older talkers and younger listeners performed equally well with younger and older ones. There are inconclusive findings in relation to speaker sex indicating that listeners may be using different strategies when judging males and females (Schötz, 2004, 2005). Such differences may be the result of lower familiarity with sociolinguistic cues for age estimation in different social groups.

One such group of speakers that listeners may have lower familiarity with is foreign-accented speakers. Non-native speech is often categorized by segmental and suprasegmental

deviation from native speaker norms, including differences in F0 and lower speech rates (Kang, 2010; Kang, Rubin, & Pickering, 2010) and imperfect acquisition of sociolinguistic patterns of variation (Adamson & Regan, 1991). This sort of variation in acquisition of native speaker norms may make assignment of social characteristics more difficult for non-native speakers; and it has been shown that listeners are more accurate with accent identification in native speakers than non-native ones (self-reference omitted).

The attested links between F0 and perceived speaker age may have implications where cross-linguistic F0 differences are concerned. There is a growing literature documenting the F0 profiles of different languages, and significant differences have been found for both typologically distinct (English and Japanese in Yamazawa & Hollien, 1992), and closely related languages (English and German in Mennen, Schaeffler, & Docherty, 2012). Bilinguals speaking their respective languages have also exhibited dissimilar F0 profiles (simultaneous bilinguals: Japanese-English in Graham, 2014; German-French and German-Italian in Voigt, Jurafsky, & Sumner, 2016), which in turn do not match those of monolingual native speakers (consecutive bilinguals compared to monolinguals: Japanese-English in Nariai & Tanaka, 2010). Such language-specific F0 patterns may result in variation in speaker age estimation across languages.

Braun and Cerrato (1999) explored the effect of the first language (L1) on speaker age estimation. They played clips recorded by native speakers of German and Italian to two groups of listeners from these L1s. The listeners were slightly better at the familiar language than the unfamiliar one, but the effect did not reach significance. There could be several explanations for the null effect for the two European languages. First, the true difference in the two speaker groups' production of variables implicated in speaker age estimation may approach zero due to substantial linguistic and socio-cultural similarities (but note that Voigt, Jurafsky, & Sumner,

2016 found significant differences in F0 profiles of the languages). It is also possible that geographical proximity allows for frequent contact, resulting in high cross-language familiarity and accuracy in assignment to social categories. In sum, a specific combination of languages may play an important role, with speakers of more distant languages and fewer opportunities for inter-personal contact exhibiting lower accuracy.

Nagao and Kewley-Port (2005) conducted a speaker age estimation study with stimuli presented in two more distant languages, English and Japanese. For the sentence-length stimuli, they found a higher correlation in the familiar language than in the unfamiliar one. There was also a general tendency to underestimate the age of middle-aged and elderly speakers (as in Hartman, 1979; Hughes & Rhodes, 2010; Mulac & Gilles, 1996), which the authors attribute to a peer-group effect. There also seemed to be a trend for general underestimation for the unfamiliar language. At closer inspection in Nagao (2006), it is revealed that in the young group, English listeners were equally accurate in both languages, and Japanese listeners underestimated the age of Japanese speakers; in the middle-aged and elderly groups, listeners were more accurate with the familiar language, and Japanese listeners were comparatively more accurate. The author attributed the Japanese listeners' higher accuracy with middle-aged and elderly speakers to higher familiarity due to more exposure to speakers from these groups; however, the familiarity account does not explain young English listeners' higher relative accuracy with Japanese speakers as one would expect Japanese listeners to have more exposure to the English language through popular culture and foreign language learning than vice versa. In terms of speaker sex, Japanese listeners estimated English females to be younger than their male counterparts, and English listeners estimated English females to be only slightly younger than males in the young and middle-aged group; no effect of speaker sex was found for Japanese speakers.

These results suggest that there may be a difference in how listeners estimate the age of speakers in different languages. Rodrigues and Nagao (2010) studied whether such an age-related sociolinguistic effect may extend to foreign-accented speech. They played English language clips recorded by male native speakers of Arabic and English to native English listeners with more and less experience with foreign accents. The listeners were asked to rate the speaker on an accentedness scale, guess the L1, and estimate their age. They found a higher correlation between estimated and chronological age for English speakers than for Arabic speakers and a higher correlation for the more experienced listeners than for the less experienced ones for Arabic speakers but not for English speakers. This finding suggests that familiarity with accents may indeed have an effect on accuracy in speaker age estimation.

To sum up, the correlations between real and estimated speaker age in the previous multilingual studies range from 0.16 for English listeners listening to Arabic-accented English to 0.89 for Japanese listeners listening to Japanese speakers speaking Japanese (Table 1). The variation in correlations has been explained by listener familiarity with the language or accents, so that speakers are better at estimating speaker age in their native language (0.58 vs 0.30 for Germans and Italians respectively listening to German speech in Braun and Cerrato [1999] and 0.81 vs 0.70 for English and Japanese respectively listening to English speech in Nagao [2006]) or accent (English listeners are better when listening to English-accented stimuli than to Arabic-accented ones in Rodrigues and Nagao [2010]), and listeners with more experience with foreign accents are better with foreign-accented speech (listeners with more experience with foreign accents are better at Arabic-accented stimuli compared to listeners with less experience in Rodrigues and Nagao [2010]).

Table 1: Correlations between real and estimated speaker age in multilingual studies

| Study | Speaker L1 | Listener L1 | Language of stimuli | Correlation |
|---|---|---|---|---|
| Braun and Cerrato (1999) | Italian | Italian | Italian | 0.75 |
| | German | German | German | 0.58 |
| | German | Italian | German | 0.30 |
| | Italian | German | Italian | 0.75 |
| Nagao (2006) | English | English | English | 0.81 |
| | Japanese | Japanese | Japanese | 0.89 |
| | English | Japanese | English | 0.70 |
| | Japanese | English | Japanese | 0.67 |
| Rodrigues and Nagao (2010) | English | English | English | 0.39 |
| | Arabic | English | English | 0.16 |

Still several questions relating to cross-linguistic speaker age estimation require elucidating. First, Rodrigues and Nagao (2010) employed native listeners, and to our knowledge, non-native listeners' speaker age estimation in second language (L2) -accented speech has not been tested. Second, previous studies often used correlations between chronological and estimated speaker age despite known issues with using correlations: a consistent under- or overestimation may result in a higher correlation despite low accuracy (Moyse, 2014). Finally,

acoustic correlates of estimated speaker age in speaker age estimation across languages have not been addressed in previous studies.

The current paper extended the study of cross-linguistic speaker age estimation to non-native speakers and listeners. It employed three measures to assess speaker age estimation: correlations between real and estimated speaker age, models of estimated speaker age, and models of speaker age estimation accuracy. We also explored the effects of acoustic (mean F0, maximum F0, minimum F0, F0 range, and rate of speech) and L2 acquisition history (speaker age of onset, age of arrival, and length of residence) measures on age estimation.

We are reporting on two perception studies in which native and non-native listeners estimated the age of native and non-native speakers of English. Models of estimated speaker age and speaker age estimation accuracy address the limitation of correlations, and using correlations allows for a comparison with previous studies. Using these different measures, we expect to find a continuum of increasing age estimation accuracy with growing speaker-listener variety familiarity. We also predict that variation in age estimation accuracy will be partially explained through acoustic and L2 acquisition differences: speakers with higher F0 and speech rate will be perceived younger, and native listeners will be more accurate with non-native speakers with a younger age of onset.

## 2. Experiment 1: English listeners

### *2.1 Method*

The audio stimuli were 40 clips of 20 England English speakers and 20 Japanese L1 speakers reading the 'Please call Stella' passage in English, retrieved from the Speech Accent Archive (Weinberger, 2015). For each language group, we aimed to have four speakers for each age category: 20s, 30s, 40s, 50s, and 60s, but this was not possible with Speech Accent Archive recordings. Table 2 shows the numbers of recordings per age group and L1. Audio-recordings of 11 English females, 9 English males, 14 Japanese females, and 6 Japanese males were used. The following English acquisition history information is available for the Japanese L1 speakers: age of English onset in years (AoO; range 10-14, mean 12.45), length of English residence in years (LoR; range 0-30, mean 10.41), and age of arrival in years (AoA, calculated by subtracting LoR from chronological speaker age; range 15-46, mean 18.34). The Japanese speakers' degree of foreign accentedness was quantified in a post hoc perception experiment with a group of 10 native English listeners who had not participated in other parts of the study and were naive to its purpose. These listeners rated the Japanese speakers' voices on a 1-7 Likert scale from 'no accent' to 'very strong accent'. An average accent strength estimate was calculated for each Japanese speaker: speaker means ranged from 2 to 6.2 on the accentedness scale (mean 3.8).

Table 2: Number of recordings used as stimuli, by age and first language of the recorded speaker

| Age group | English L1 | Japanese L1 |
|---|---|---|
| 18-29 | 7 (4 males) | 7 (4 males) |
| 30-39 | 4 (1 male) | 3 (1 male) |
| 40-49 | 4 (2 males) | 5 (1 male) |

| | | |
|---|---|---|
| 50-59 | 3 (2 males) | 4 (0 males) |
| 60+ | 2 (0 males) | 1 (0 males) |

The audio files in the Speech Accent Archive are available in the mp3 format (128 kbps); they were converted to wav for F0 tracking. Such acoustic analysis of re-converted compressed files is justified as measures of F0 stay reliable with reported mean errors below 2% for 56-320 kbps (Fuchs & Maxwell, 2016). Following Mennen, Schaeffler, and Docherty (2012), several long-term distributional measures were obtained for F0 using Praat (Boersma & Weenink, 2009): mean F0, maximum F0, minimum F0, and F0 range (calculated by subtracting minimum F0 from maximum F0). The standard Praat settings were used: namely, the autocorrelation method, the pitch floor of 75 Hz and the pitch ceiling of 500 Hz. Manual corrections were used to increase reliability (e.g., creaky voice was excluded). Mean duration of clips (which is a reflection of speech rate) was 21.6 and 27.3 seconds for the English and Japanese speakers respectively. Table 3 contains these mean F0 and duration measures and their standard deviations (SD).

Table 3. Mean (SD) F0 measurements in Hz and duration in seconds.

| Speakers | mean F0 | max F0 | min F0 | F0 range | Duration |
|---|---|---|---|---|---|
| English females | 203 (23) | 454 (71) | 86 (28) | 368 (70) | 21.9 (2.6) |
| English males | 145 (71) | 430 (109) | 83 (15) | 347 (105) | 24.2 (3.6) |
| Japanese females | 199 (20) | 360 (67) | 120 (7) | 240 (69) | 28.6 (4.0) |

| | | | | | |
|---|---|---|---|---|---|
| Japanese males | 122 (119) | 183 (81) | 81 (10) | 103 (88) | 21.2 (1.8) |

Thirty-six British English L1 participants listened to the audio stimuli via Sony MDR ZX310 headphones and, after each stimulus finished, were asked to estimate that speaker's age (once per speaker). The experiment was run in a laboratory setting using PsychoPy version 1.83.03 (Peirce, 2009). Participants' age estimates and reaction times (from the end of audio stimulus presentation) were recorded. The order of stimuli was randomised for each participant. The ages of these listeners ranged from 19 to 57, with a mean of 23.5 years. Thirty listeners identified themselves as female, and 6 as male. All of them reported English or British English as one of their native languages.

## *2.2 Data analysis*

We calculated correlations between real and estimated speaker age for comparison with previous studies and fit regression models for predicting estimated speaker age from real speaker age and for predicting the absolute difference between real and estimated speaker age for an analysis of accuracy.

We fit several linear mixed effects models (Baayen, Davidson, & Bates, 2008) to the data in R (R Core Team, 2017). Fixed effects are the independent variables whose effect is being studied (e.g., speaker L1); random effects are sources of variance in a subset of general population being used (e.g., participants); random slopes allow for fixed effects to vary in relation to a given random effect (e.g., the effect of speaker L1 may vary from one listener to another due to individual differences; Barr et al. 2013). In each case, we started with a more saturated model as specified in the Results section, and the model was pruned to keep only

significant effects (with $p < 0.05$) and effects significantly improving model fit (as identified by an analysis of variance comparing a current model with the previous one that excludes one non-significant predictor and repeating this process recursively). Certain random slopes were excluded for the benefit of model convergence (Matuschek et al., 2017). In these analyses, age, listener age, speaker age, AoA, AoO, LoR, age estimates, and reaction times were log-transformed because of skewness of data.[1]

In the results tables female English speakers were used as the baseline (females in the Japanese only models). The *estimate* and the *standard error* columns in the tables give us the predicted estimated speaker age and standard error for a level respectively. To calculate the predicted estimated speaker age for a different level, the respective value in the *estimate* column is added or subtracted. The last column indicates whether the observed differences are significant.

## *2.3 Results*

The Pearson's product moment correlation between real age and estimated age for English and Japanese L1 speakers is 0.64 ($t = 22.199$, $df = 718$, p-value $< 0.001$) and 0.37 ($t = 10.523$, $df = 717$, p-value $< 0.001$) respectively. The distribution of correlation coefficients from 1000 random permutations of the English listeners' data is centered on 0, which shows that the observed correlation coefficients of 0.64 and 0.37 are significantly different from 0. The largest absolute difference between correlation coefficients in these 1000 permutations is 0.19, which is evidently smaller than the difference between the observed coefficients (0.64 - 0.37 = 0.27). This shows

[1] Log-transforming reduced $G_1$ (Joanes & Gill, 1998, 184) for speakers' ages (from 0.45 to -0.10), LoR (from 0.54 to -0.39), AoA (from 0.36 to -0.07), English listeners' ages (from 2.97 to 2.42), English listeners' reaction times (from 5.64 to 1.16), English listeners' estimates (from 0.69 to -0.10), and Japanese listeners' estimates (from 0.58 to -0.13). Log-transforming increased $G_1$ for Japanese listeners' ages (from -0.38 to -0.96) and AoO (from -1.00 to -1.24).

that the difference between the two observed correlation coefficients is significantly different from 0. The distribution of correlation coefficients from 1000 bootstrap samples of the English listeners' data is centered on the observed values. The two distributions of bootstrap correlation coefficients (for English L1 speakers on the one hand and Japanese L1 speakers on the other) do not overlap, which supports the analysis that the two observed correlation coefficients are significantly different from each other.

Linear models for English and Japanese speakers can be seen in Figure 1. The red line below the black line indicates that Japanese speakers were perceived to be younger than their English counterparts. It can be seen that the oldest Japanese L1 speaker falls outside of the general pattern and carries a strong effect. If the three 60+ speakers are excluded, the significant positive correlations and relative position of regression lines hold. The correlation between real age and estimated age for English and Japanese L1 speakers becomes 0.54 ($t = 16.479$, $df = 646$, p-value $< 0.001$) and 0.48 ($t = 14.099$, $df = 681$, p-value $< 0.001$) respectively (Figure 2). The distribution of correlation coefficients from 1000 random permutations of the English listeners' data for speakers aged 60 and under is centered on 0, which shows that the observed correlation coefficients of 0.54 and 0.48 are significantly different from 0. The largest absolute difference between correlation coefficients in these 1000 permutations is 0.29, which is larger than the difference between the observed coefficients (0.54 - 0.48 = 0.06). The distribution of correlation coefficients from 1000 bootstrap samples of the English listeners' data is centered on the observed values. Although the two distributions of bootstrap correlation coefficients (for English L1 speakers under 60 on the one hand and Japanese L1 speakers under 60 on the other) overlap somewhat, due to the smaller difference between observed correlation coefficients, the fact that

the respective distributions of permutation-based correlation coefficients center on the observed values strengthens these observed values.

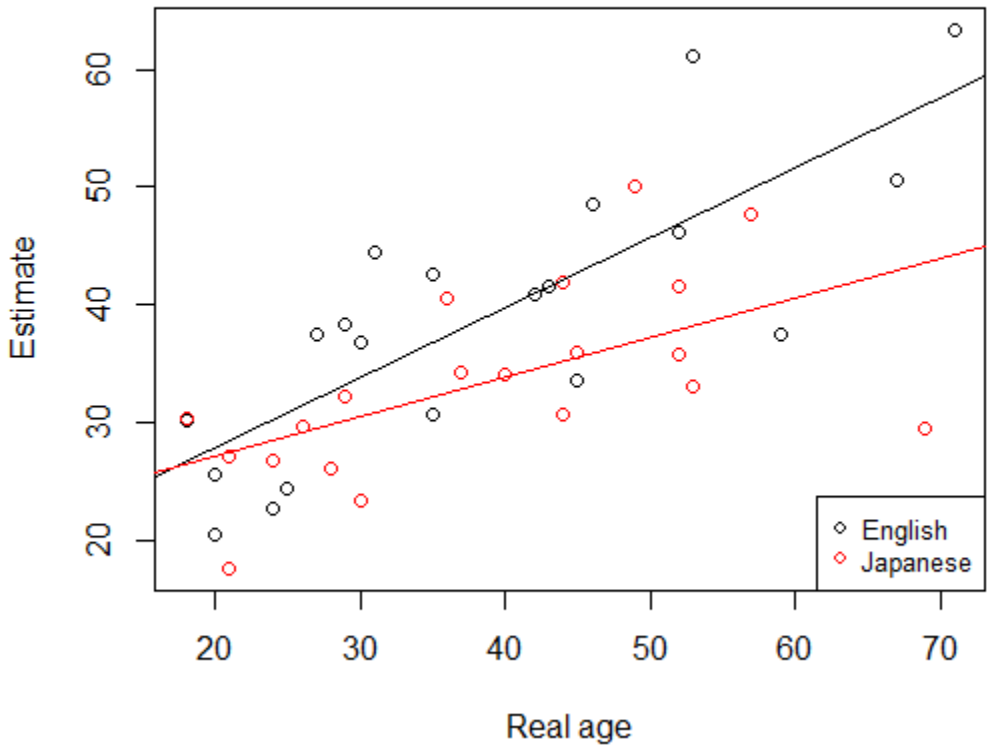


Figure 1: Speaker real ages plotted against their mean age estimate

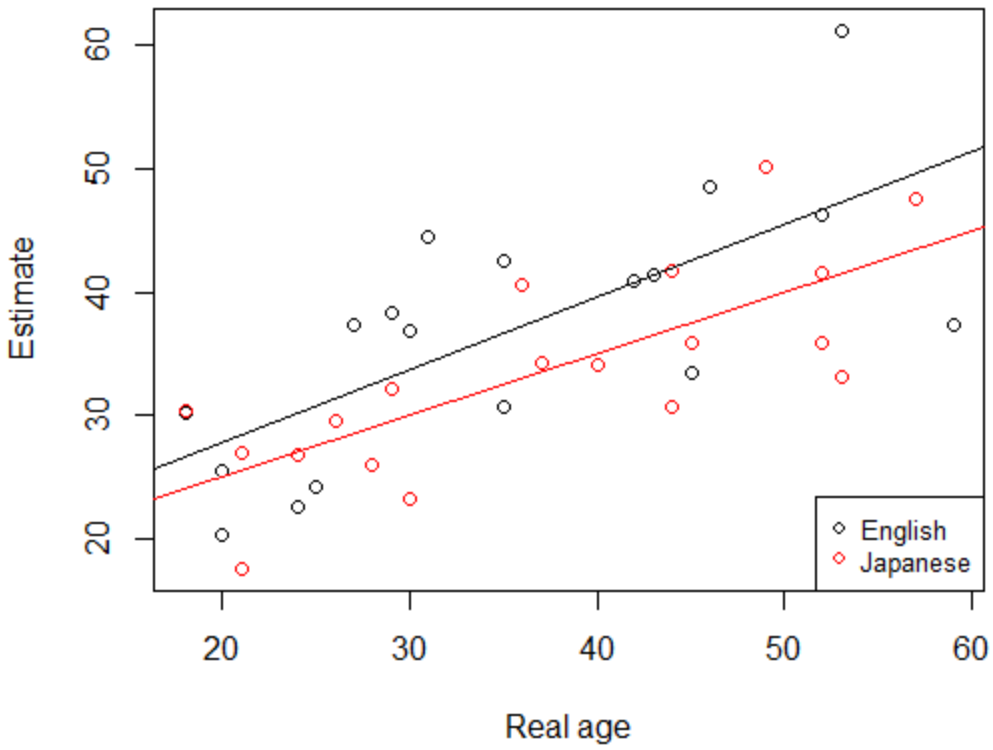


Figure 2: Speaker real ages plotted against their mean age estimate (speakers 60+ excluded)

Because correlation may not be a good reflection of accuracy, we have run another set of statistical tests with all the 40 speakers conceptualizing accuracy as the absolute difference between real and estimated speaker age.

For a model of accuracy, we used listener age, speaker sex, L1, and age, as well as two two-way interactions between speaker and listener age and between speaker sex and L1 as fixed effects; speaker and listener as random effects; and speaker sex, L1, age, and an interaction between speaker sex and L1 as random slopes for listener (formula: accuracy ~ speaker_age*listener_age + speaker_sex*speaker_L1 + (1 + speaker_L1*speaker_sex +

speaker_age|listener) + (1|speaker)). Speaker_age was excluded as random slope for the benefit of model convergence. In the end, none of the independent variables was a significant predictor of accuracy, but speaker age, sex, and L1 all individually significantly improved model fit (Table 4). The model revealed a trend (p=0.082) for listeners to be less accurate with Japanese L1 speakers.

Table 4: Summary for model of accuracy

| | Estimate | Std. Error | df | t value | Pr(>\|t\|) | Sign. |
|---|---|---|---|---|---|---|
| (Intercept) | -0.080 | 0.215 | 40.040 | -0.372 | 0.712 | |
| speaker_age | 0.089 | 0.057 | 39.950 | 1.565 | 0.126 | |
| speaker_sex--male | 0.022 | 0.046 | 44.300 | 0.470 | 0.641 | |
| speaker_L1--Japanese | 0.075 | 0.042 | 43.510 | 1.783 | 0.082 | |

To create a model for predicting estimated speaker age from real speaker age, another mixed-effects model was fit to the data with the age estimate as the dependent variable; listener age, speaker sex, L1, and age, as well as two two-way interactions between speaker and listener age and between speaker sex and L1 as fixed effects; speaker and listener as random effects; and speaker sex, L1, age, and an interaction between speaker sex and L1 as random slopes for listener (formula: estimate ~ speaker_age*listener_age + speaker_sex*speaker_L1 + (1 + speaker_L1*speaker_sex + speaker_age|listener) + (1|speaker)). The final model is presented in Table 5. As expected, speaker age was a significant predictor of estimated age, supporting

previous claims that listeners are fairly accurate at estimating speakers' age. Speaker sex and L1 were also significant predictors, such that male speakers were estimated to be older than females and Japanese L1 speakers were estimated to be younger than English L1 speakers.

Table 5: Summary for model of age estimate (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$)

| | Estimate | Std. Error | df | t value | Pr(>\|t\|) | Sign. |
|---|---|---|---|---|---|---|
| (Intercept) | 1.460 | 0.288 | 41.530 | 5.073 | 0.000 | |
| speaker_age | 0.580 | 0.077 | 41.710 | 7.582 | 0.000 | *** |
| speaker_sex--male | 0.152 | 0.061 | 43.820 | 2.498 | 0.016 | * |
| speaker_L1--Japanese | -0.132 | 0.058 | 48.980 | -2.290 | 0.026 | * |

To assess what factors affect speaker age estimation in foreign-accented speakers, we fit a linear mixed effects model to Japanese speaker data with the age estimate as the dependent variable; clip duration, speaker age, LoR, AoO, AoA, and degree of foreign accentedness, as well as four two-way interactions between speaker sex and mean F0, max F0, min F0, and F0 range as fixed effects and random slopes for listener; and speaker and listener as random effects (formula: estimate ~ duration + speaker_age + LoR + AoO + AoA + accent + speaker_sex*meanF0 + speaker_sex*maxF0 + speaker_sex*minF0 + speaker_sex*F0range + (1 + duration + speaker_age + LoR + AoO + AoA + accent + speaker_sex*meanF0 + speaker_sex*maxF0 + speaker_sex*minF0 + speaker_sex*F0range|listener) + (1|speaker)). We find a significant effect of AoA such that speakers with a higher age of arrival are estimated to be

older (Table 6). Next, there is a significant effect of mean F0 such that speakers with a higher F0 are estimated to be younger. Finally, a significant effect of speaker sex suggests that male speakers are estimated to be younger than their female counterparts of the same AoA and mean F0.

Table 6: Summary for model of age estimate in Japanese L1 speakers (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$)

| | Estimate | Std. Error | df | t value | Pr(>\|t\|) | Sign. |
|---|---|---|---|---|---|---|
| (Intercept) | 2.023 | 0.380 | 20.078 | 5.319 | 0.000 | |
| speaker_sex--male | -0.414 | 0.166 | 20.100 | -2.499 | 0.021 | * |
| mean F0 | -0.215 | 0.076 | 19.878 | -2.826 | 0.010 | * |
| AoA | 0.473 | 0.112 | 19.873 | 4.230 | 0.000 | *** |

To investigate the factors affecting reaction time, we fit a linear mixed effects model with reaction time as the dependent variable; listener age, speaker sex, L1, age, and order in the experiment, as well as two two-way interactions between speaker and listener age and between speaker sex and L1 as fixed effects; speaker and listener as random effects; and speaker sex, L1, age, order, and an interaction between speaker sex and L1 as random slopes for listener (formula: reaction_time~ speaker_age*listener_age + speaker_sex*speaker_L1 + order + (1 + speaker_L1*speaker_sex + speaker_age + order|listener) + (1|speaker)). As is often the case in reaction time experiments, we find a significant effect of order: participants get faster in the

course of the experiment (Table 7). There is also a significant effect of speaker L1 such that participants are slower reacting to Japanese females. This is mediated by a trend towards an interaction with speaker sex (p=0.078) which significantly improved model fit: listeners are faster reacting to Japanese males than Japanese females.

Table 7: Summary for model of reaction time (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$)

| | Estimate | Std. Error | df | t value | Pr(>\|t\|) | Sign |
|---|---|---|---|---|---|---|
| (Intercept) | 1.670 | 0.049 | 38.450 | 34.326 | 0.000 | |
| speaker_sex--male | 0.064 | 0.047 | 35.110 | 1.374 | 0.178 | |
| speaker_L1--Japanese | 0.131 | 0.048 | 38.180 | 2.716 | 0.010 | ** |
| order | -0.008 | 0.001 | 34.620 | -6.256 | 0.000 | *** |
| speaker_sex--male: speaker_L1--Japanese | -0.122 | 0.067 | 40.370 | -1.809 | 0.078 | |

## *2.4 Interim discussion*

In the first experiment involving English listeners, similar to previous research, we find that listeners are able to estimate speakers' age from speech. We find significant positive correlations between real speaker age and estimated age which are slightly lower than those reported by Nagao and Kewley-Port (2005) and higher than Rodrigues and Nagao (2010). Moreover, speaker age and AoA emerge as significant predictors of estimated age. There is a correlation of 0.74

between AoA and speaker age, so the significance of AoA is most probably a reflection of speaker age. In an additional linear regression model predicting AoA from speaker age run post-hoc to explore the relationship between these two variables, we find a shallow slope, meaning that there is less variation in the AoA and the effect size on estimated age is smaller. We assume that AoA was significant instead of speaker age because it captures an additional quality related to L2 acquisition, despite other L2-related factors such as LoR, AoO, degree of accentedness, and duration not being significant.

We also find an effect of speaker L1: better age estimation for English L1 speakers (correlation of 0.64) than Japanese L1 speakers (0.37). Reflective of the difference in correlations between speaker L1s, there is a trend for listeners to be more accurate with English speakers than Japanese speakers in the model of accuracy. In the model of reaction time, slower reactions to Japanese speakers may be revealing of lower familiarity with the Japanese accent and also comparative difficulty in age estimation of accented speech. Taken together, these findings speak to poorer age estimation in accented speech, which is also in line with previous findings.

Visual analysis of the relative position of the regression lines in Figures 1 and 2 suggests that Japanese speakers are estimated to be younger than English speakers. The significance of speaker L1 in the model of estimated speaker age supports this observation statistically. Speaker sex also surfaces as a significant predictor in the models of estimated speaker age as a main effect or in an interaction, suggesting that female speakers are perceived to be younger than males. An exploration of different predictors in the model of estimated age in Japanese speakers suggests that listeners rely on mean F0 when estimating age in foreign-accented speech and associate higher F0s with younger age (see also Schötz, 2007), though this can not be the only

effect responsible for the underestimation of Japanese speakers' age as their mean F0 was not higher than that of English speakers (Table 3).

# 3. Experiment 2: Japanese listeners

## *3.1 Method*

Twenty three Japanese L1 participants listened to the same audio stimuli used in Study 1 and were asked to estimate each speaker's age (once per speaker). The experiment was run as an online survey, in four different versions. The versions differed only in the order of stimuli presentation - each version used a pre-determined, but random, order of stimuli. The participants could complete the survey in their own time on their own device, and could enter their answer before the end of the stimulus due to technical limitations of the online presentation software. In this sense there was more heterogeneity in experiment procedure in Study 2 compared to Study 1, but it allowed us to recruit participants that we otherwise would have not been able to. Participants' age estimates were recorded, but reaction times were not, as recording reliable reaction times in this mode of online presentation was not feasible. The ages of these Japanese listeners ranged from 19 to 70, with a mean of 42.9 years. Fourteen listeners identified themselves as female, and 9 as male. All of them reported Japanese as their native language. We did not measure or elicit the English proficiency level of these listeners, but we assume a relatively high one as the experiment instructions were in English and the participants were recruited via the friend-of-friend method and through English-language notices. Because of these methodological differences between Studies 1 and 2, we refrain from performing statistical modeling on both groups at the same time, but we compare the results to each other and the previous studies in the General Discussion.

### *3.2 Results*

An analysis similar to the one performed in Study 1 was done using the Japanese listener data. The Pearson's product moment correlation between real age and estimated age for English and Japanese L1 speakers is 0.44 ($t = 10.606$, $df = 458$, p-value $< 0.001$) and 0.45 ($t = 10.874$, $df = 458$, p-value $< 0.001$) respectively. The distribution of correlation coefficients from 1000 random permutations of the Japanese listeners' data is centered on 0, as is the distribution of differences between the correlation coefficients in each permutation. This shows that the observed values (0.44 and 0.45) are significantly different from 0, and that the difference between them (0.01) is not significantly different from 0.

As in Study 1, the significant positive correlations hold if the three 60+ speakers are excluded. The correlation between real age and estimated age for English and Japanese L1 speakers becomes 0.38 ($t = 8.455$, $df = 412$, p-value $< 0.001$) and 0.58 ($t = 14.988$, $df = 435$, p-value $< 0.001$) respectively. The distribution of correlation coefficients from 1000 random permutations of the Japanese listeners' data for speakers aged 60 and under is centered on 0, and the distribution of correlation coefficients from 1000 bootstrap samples is centered on the observed values.

Linear models for English and Japanese speakers can be seen in Figure 3. Once again visual analysis of the figure suggests that Japanese speakers are estimated to be younger than English speakers.

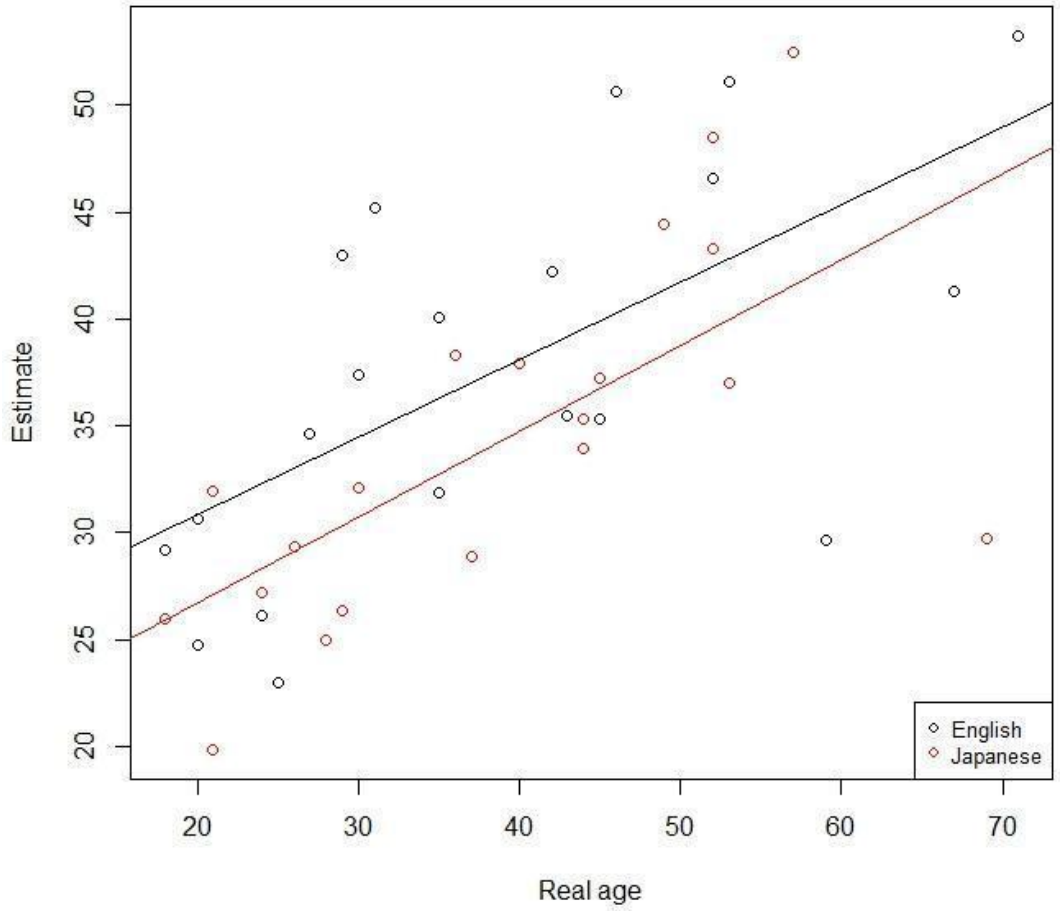


Figure 3: Speaker real ages plotted against their mean age estimate

To estimate the accuracy of speaker age estimation, we fit a linear mixed effects model with the absolute difference between real and estimated speaker age as the dependent variable; listener age, speaker sex, L1, and age, as well as two two-way interactions between speaker and listener age and between speaker sex and L1 as fixed effects; speaker, listener, and experiment version as random effects; and speaker sex, L1, age, and an interaction between speaker sex and L1 as random slopes for listener (formula: accuracy ~ speaker_age*listener_age + speaker_sex*speaker_L1 + (1 + speaker_L1*speaker_sex + speaker_age|listener) + (1|speaker) +

(1|version)). In the end, none of the independent variables was a significant predictor of accuracy.

To create a model for predicting estimated age from real speaker age, another mixed-effects model was fit to the data with the age estimate as the dependent variable; listener age, speaker sex, L1, and age, as well as two two-way interactions between speaker and listener age and between speaker sex and L1 as fixed effects; speaker, listener, and experiment version as random effects; and speaker sex, L1, age, and an interaction between speaker sex and L1 as random slopes for listener (formula: estimate ~ speaker_age*listener_age + speaker_sex*speaker_L1 + (1 + speaker_L1*speaker_sex + speaker_age|listener) + (1|speaker) + (1|version)). The final model is presented in Table 8. Speaker age was a significant predictor of estimated age, supporting previous findings that listeners are fairly accurate at estimating speakers' age, extending this claim to non-native listeners. Speaker sex was also significant, so that male speakers were estimated to be older than females. No effect of speaker L1 was found for Japanese listeners.

Table 8: Summary for model of age estimate (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$)

| | Estimate | Std. Error | df | t value | Pr(>\|t\|) | Sign. |
|---|---|---|---|---|---|---|
| (Intercept) | 1.778 | 0.293 | 45.130 | 6.070 | 0.000 | |
| speaker_age | 0.478 | 0.079 | 45.450 | 6.080 | 0.000 | *** |
| speaker_sex--male | 0.120 | 0.060 | 44.460 | 1.988 | 0.053 | |

To assess what factors affect age estimation in Japanese L1 speakers, we fit another linear mixed effects model to Japanese speaker data only, with estimate as the dependent variable; clip duration, speaker age, LoR, AoO, AoA, and degree of accent, as well as four two-way interactions between speaker sex and mean F0, max F0, min F0, and F0 range as fixed effects and random slopes for listener; and speaker, listener, and experiment version as random effects (formula: estimate ~ duration + speaker_age + LoR + AoO + AoA + accent + speaker_sex*meanF0 + speaker_sex*maxF0 + speaker_sex*minF0 + speaker_sex*F0range + (1 + duration + speaker_age + LoR + AoO + AoA + accent + speaker_sex*meanF0 + speaker_sex*maxF0 + speaker_sex*minF0 + speaker_sex*F0range|listener) + (1|speaker) + (1|version)). We find a significant effect of AoA and LoR such that speakers with a higher age of arrival and longer length of residence are estimated to be older. Next, there is a significant effect of mean F0 such that speakers with a higher F0 are estimated to be younger.

Table 9: Summary for model of age estimate in Japanese L1 speakers (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$)

| | Estimate | Std. Error | df | t value | Pr(>\|t\|) | |
|---|---|---|---|---|---|---|
| (Intercept) | 2.016 | 0.349 | 21.041 | 5.768 | 0.000 | |
| meanF0 | -0.080 | 0.036 | 23.592 | -2.214 | 0.037 | * |
| LoR | 0.142 | 0.030 | 23.931 | 4.754 | 0.000 | *** |
| AoA | 0.366 | 0.101 | 20.733 | 3.616 | 0.002 | ** |

### *3.3 Interim discussion*

The correlations between real and estimated speaker age in this experiment are between the values reported by Nagao and Kewley-Port (2005) and Rodrigues and Nagao (2010); however, there is no longer an advantage for English L1 speakers over Japanese L1 speakers (or even a slight disadvantage if the 60+ speakers are excluded), suggesting that Japanese listeners are equally good at estimating the age of both English and Japanese speakers. Supportive of this is also the finding of no significant effect of speaker L1 in the accuracy and estimated age models.

Other factors that reach significance in this study are speaker sex, mean F0, AoA, and LoR. Male speakers are estimated to be older than their female counterparts. The effect of mean F0 is in the same direction: speakers with a higher mean F0 are estimated to be younger. The significant effect of AoA is most probably a reflection of the effect of speaker age which was found to be significant for English L1 listeners. The additional effect of LoR, which was not found in the English listener model, may be reflective of the speakers' acquisition of L2 norms, such as the F0 contour, affecting speaker age estimation.

## 4. General discussion

This study adds to previous literature on the effect of L1 in speaker age estimation. We found similar trends in variation in the current study: English listeners exhibited the highest correlation of 0.64 with English speakers and the lowest correlation of all when listening to Japanese-accented English (0.37). Japanese listeners fell between these values, performing equally well with English- and Japanese-accented English (0.44 and 0.45). If we use Nagao (2006) to make an extrapolation about these Japanese listeners listening to the Japanese speakers speaking Japanese, we can expect to see an even higher correlation. This continuum reflects the relative familiarity

of the listeners with accents. English listeners presumably have the most experience with English accents in their native language and are less familiar with Japanese-accented English. Japanese listeners are all second language speakers of English and they are familiar with both the English and the Japanese accent in English through everyday experiences. The similarity of correlations in Japanese listeners' estimations of English- and Japanese-accented speech is reminiscent of the matched interlanguage speech intelligibility benefit found in speech intelligibility studies, which show that non-native listeners find non-native speakers of the same L1 background as intelligible as native speakers (Bent & Bradlow, 2003).

The lower correlation found for Japanese listeners listening to Japanese speakers in English (0.45) in comparison to Japanese listeners listening to Japanese speakers in Japanese (0.89) in Nagao (2006) suggests that some age-related information is lost in such cross-linguistic estimations. The within-speaker F0 differences across languages in the same bilingual speaker as found by, for example, Graham (2014) could be one explanation for such behaviour. If Japanese L1 speakers speak with a lower F0 on average in English than in Japanese and listeners rely on speaker F0 as their age estimation cue, this would result in lower age estimation accuracy for English language stimuli. Designing a study in which bilingual listeners estimate the age of bilingual speakers across both languages would allow to explore this further.

Some differences in correlations across studies could also be explained through varying methodology. We employed longer stimuli which should provide more information to base the judgment on. We find correlations lower than those for passage-length stimuli in Ryan & Capadano (1978), but higher than those found in Rodrigues and Nagao (2010), the most comparable study that also used foreign-accented speech and a 20-word extract from the 'Please call Stella' passage.

Our two studies with English and Japanese listeners did not find a significant relationship between speaker age estimation accuracy and speaker L1 (despite a trend in English listeners), demonstrating that despite differences in correlations, there may be similar accuracy due to a consistent under- or overestimation (as suggested by Moyse, 2014). The trend towards being less accurate with Japanese speakers by English listeners and English listeners being significantly slower when estimating Japanese speakers' age point in the same direction as the differences in correlations: listeners least familiar with the accent were less accurate and slower in speaker age estimation.

Modeling estimated age as a function of real speaker age revealed a significant effect of speaker sex and L1 for English listeners and a trend for speaker sex for Japanese listeners. Both groups of listeners estimated male speakers to be significantly older than their female counterparts. These results are in line with Nagao (2006), who found a similar effect of speaker sex for English females. As we argue for the speaker L1 effect below, there may be a relationship with F0. Women generally speak with a higher F0 than men because of physiological differences, but there are also cultural effects such that an even higher F0 is found in Japanese females (e.g., Loveday, 1981). This may result in females being perceived to be younger than males. Certainly, we acknowledge the unbalanced sex distribution in our sample with a larger number of females than males, especially in the older age groups. The findings of this research should be considered in light of this limitation.

The speaker L1 effect means that Japanese speakers were estimated to be younger than the English speakers by the English listeners, but not by the Japanese listeners. For comparison, in Nagao and Kewley-Port (2005) there was a trend towards underestimation of age in the unfamiliar language. We argue that this is another manifestation of listener familiarity: in the

current study, English listeners underestimated the age of Japanese speakers; Japanese listeners were equally familiar with English- and Japanese-accented speech, so accent did not affect their age estimation as they were able to adjust their schemata accordingly for both familiar accents.

The question remains what it is that marks age and requires familiarization. Two of the variables that have been suggested to affect estimated age are rate of speech and F0 (e.g., Schötz, 2007). In the models of estimated age for Japanese speakers only, an effect of mean F0 was found for both English and Japanese listeners such that speakers with a higher F0 were estimated to be younger, suggesting that this relationship extends to foreign-accented speech and non-native listeners. Additionally, the absence of familiar language-specific sociolinguistic variation in a foreign language or imperfect acquisition of such language-specific sociolinguistic rules which could be a speaker age cue (Drager, 2010; Walker, 2007) would also result in a language effect in age estimation. We acknowledge the limitation that global F0 measures used in this paper are very rough measures for passage-length stimuli and may not be the only acoustic dimension capturing age differences as demonstrated by previous research (Schötz, 2007); we hope that future research will help to identify other contributing factors in cross-linguistic age estimation.

For this study, we did not collect social network information about the listeners and, therefore, the familiarity argument is based on our assumptions about listener familiarity with these accents. Additionally, we acknowledge the differences in data collection procedure in the two studies which prevented us from analysing the differences between the two groups of listeners statistically and made the two studies less readily comparable. Future perception research will benefit from collecting more information about listeners and correlating detailed measures of exposure to different accents with accuracy in speaker age estimation. In turn,

production studies could investigate acoustic characteristics in speakers of different languages at different ages to determine whether languages generally differ in their age-related acoustic features. In the present study, we examined speaker age and F0, but it is well known that older and younger speakers differ with regard to other segmental and suprasegmental features (e.g., Drager, 2010; Walker, 2007). Just as F0 appears to differ between languages (e.g., Yamazawa & Hollien, 1992), so might these other potential cues for speaker age differ between languages. If this is the case, speakers of one language could not reliably estimate the ages of speakers of other language by the language-specific sociolinguistic patterns found in their own language.

The significant effect of speaker L1 and the differences in age estimation by English and Japanese listeners highlight that even such a seemingly universal phenomenon as age may be expressed and perceived differently by people from different language backgrounds and of varying familiarity with languages and accents. This further supports the previous studies that show a connection between speaker age and sociolinguistic features, reflecting that age is expressed both physiologically and socio-culturally. The practical implications of this study include our need for awareness of such differences when speaker age estimation occurs in real life, for example when relying on amateur estimates of a speaker's age for forensic purposes in cases of ear-witnesses. Ear-witness identifications are not entirely reliable even when witnesses know the speaker (e.g., Foulkes & Barron, 2000), so heightened caution with unfamiliar accents is recommended.

## 5. Conclusion

Age determines many aspects of social interaction across cultures. Children may be pampered in one society and marginal in another, old people may be revered for their wisdom in one society

and pitied for their reduced physical capabilities in another - the fact that a person's age affects how they are treated is common to all of them. This, of course, includes language: age-graded linguistic forms, Japanese honorifics, and politeness in speech all encode the speaker's and addressee's age, among other characteristics.

This study addressed speaker age estimation across languages and the effect of speaker and listener L1s on age estimation. This is one of the few studies that considered age estimation in bilingual speakers and listeners, and the first study that we know of that included listeners and speakers of a matching L1 with stimuli in the L2. The results suggest that listeners may not only differ in how they estimate the age of speakers in different languages but also in how they estimate the age of speakers with different foreign accents. As we do not expect physiological differences between speakers of different languages to explain all acoustic differences between speakers of different languages, we explained this through listener variation in familiarity with how different languages may mark age sociolinguistically.

## References

Adamson, H. D., & Regan, V. M. (1991). The acquisition of community speech norms by Asian immigrants learning English as a second language. A preliminary study. *Studies in Second Language Acquisition, 13*(1), 1-22.

Baayen, R.H, Davidson, D. J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Barr, D.J., Levy, R., Scheepers, C.C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Benjamin, G.R. (1992). Perceptions of speaker's age in natural conversation in Japanese and English. *Language Sciences, 14*(1/2), 77-87.

Bent, T., & Bradlow, A.R. (2003). The Interlanguage Speech Intelligibility Benefit. *The Journal of the Acoustical Society of America, 114*(3), 1600.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 6.0.24). Retrieved from http://www.praat.org/

Braun, A., & Cerrato, L. (1999). Estimating speaker age across languages. *ICPhS 99,* 1369–1372.

Cerrato, L., Falcone, M., & Paolini, A. (2000). Subjective age estimation of telephonic voices. *Speech Communication, 31,* 107–112. DOI: http://dx.doi.org/10.1016/S0167-6393(99)00071-0

Decoster, W., & Debruyne, F. (1996). The ageing voice: changes in fundamental frequency, waveform stability and spectrum. *Acta oto-rhino-laryngologica belgica, 51(2),*105–112.

Drager, K. (2010). Speaker age and vowel perception. *Language and Speech, 54*(1), 99-121.

Foulkes, P., & Barron, A. (2000). Telephone speaker recognition amongst members of a close social network. *International Journal of Speech Language and the Law, 7*(2), 180-198.

Fuchs, R., & Maxwell, O. (2016). The Effects of mp3 Compression on Acoustic Measurements of Fundamental Frequency and Pitch Range. In proceedings of Speech Prosody 8, Boston, USA.

Graham, C. (2014). Fundamental Frequency Range in Japanese and English: The Case of Simultaneous Bilinguals. *Phonetica, 71*, 271-295.

Harnsberger, J. D., Brown, W. S., Shrivastav, R., & Rothman, H. (2010). Noise and tremor in the perception of vocal aging in males. *Journal of Voice, 24*, 523–530. DOI: http://dx.doi.org/10.1016/j.jvoice.2009.01.003

Harnsberger, J. D., Shrivastav, R., Brown, W. S., Rothman, H., & Hollien, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice, 22*, 58–69. DOI: http://dx.doi.org/10.1016/j.jvoice.2006.07.004

Hartman, D. E. (1979). The perceptual identity and characteristics of aging in normal male adult speakers. *Journal of Communication Disorders, 12*, 53–61.DOI: http://dx.doi.org/10.1016/0021-9924(79)90021-2

Hughes, S. M., & Rhodes, B. C. (2010). Making age assessments based on voice: The impact of the reproductive viability of the speaker. *Journal of Social, Evolutionary, and Cultural Psychology, 4*, 290–304. DOI: http://dx.doi.org/10.1037/h0099282

Hunter, E.J., Ferguson, S.H., & Newman, C.A. (2016). Listener estimations of talker age: A meta-analysis of the literature. *Logopedics Phoniatrics Vocology, 41*(3), 101-105. DOI: 10.3109/14015439.2015.1009160

Huntley, R., Hollien, H., & Shipp, T. (1987). Influences of listener characteristics on perceived age estimations. *Journal of Voice, 1*, 49–52. DOI:http://dx.doi.org/10.1016/S0892-1997(87)80024-3

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *The Statistician, 47*(1), 183–189. DOI: https://doi.org/10.1111/1467-9884.00122

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System, 38*(2), 301-315.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal, 94*(4), 554-566.

Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology, 38*, 618–625. DOI: http://dx.doi.org/10.1016/S0022-1031(02)00510-3

Linville, S.E. (2001). *Vocal aging*. San Diego: Singular Thomson Learning.

Loveday, L. (1981). Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech, 24*(1), 71-89.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305-315.

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America, 131*(3), 2249-2260.

Moyse, E. (2014). Age Estimation from Faces and Voices: A Review. *Psychologica Belgica, 54*(3), 255-265. DOI: http://dx.doi.org/10.5334/pb.aq

Moyse, E., Beaufort, A., & Brédart, S. (2014). Evidence for an own-age bias in age estimation from voices in older persons. *European Journal of Ageing, 11*, 241-247.

Mulac, A. & Gilles, H. (1996). 'You're only as old as you sound': Perceived vocal age and social meanings. *Health Communication, 8,* 199–215. DOI: http://dx.doi.org/10.1207/s15327027hc0803_2

Nagao, K. (2006). *Cross-language study of age perception*. (PhD), Indiana University.

Nagao, K. & Kewley-Port, D. (2005). *The Effect of Language Familiarity on Age Perception.* Poster presented at the International Research Conference on Aging and Speech Communication, Bloomington, USA.

Nariai, T., & Tanaka, K. (2010). A Study of Pitch Patterns of Sentence Utterances by Japanese Speakers of English in Comparison with Native Speakers of English. Poster presented at Interspeech Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology, Tokyo, Japan.

Peirce, J.W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2(10)*. DOI: dx.doi.org/10.3389/neuro.11.010.2008

Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of Speech Hearing Research, 9*, 273–277.

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders, 16*(3), 217-226.

Ramig, L. A. (1986). Aging speech: Physiological and sociological aspects. Language and Communication, 6, 25-34.

Ramig, L. A., & Ringel, R. L. (1983). Effects of physiological aging on selected acoustic characteristics of voice. *Journal of Speech, Language, and Hearing Research, 26*(1), 22-30.

Rodrigues, P., & Nagao, K. (2010). *Effects of listener experience with foreign accent on perception of accentedness and speaker age.* Poster presented at the 159th Meeting of the Acoustical Society of America, Baltimore, USA.

Ryan, W. J. & Burk, K. W. (1974). Perceptual and acoustic correlates of aging in the speech of males. *Journal of Communication Disorders, 7*, 181–192. DOI: http://dx.doi.org/10.1016/0021-9924(74)90030-6

Ryan, E. B., & Capadano, H. L. (1978). Age perceptions and evaluative reactions toward adult speakers. *Journal of Gerontology, 33*, 98–102. DOI: http://dx.doi.org/10.1093/geronj/33.1.98

Schötz, S. (2004). *The role of F0 and duration in perception of female and male speaker age.* Paper presented at Speech Prosody, Nara, Japan.

Schötz, S. (2005). Effects of stimulus duration and type on perception of female and male speaker age. In *Proceedings of Fonetik* (pp. 87-90), Department of Linguistics, Göthenburg University.

Schötz, S. (2007). Acoustic Analysis of Adult Speaker Age. In C. Müller (Ed.), *Speaker Classification I: Fundamentals, Features, and Methods* (pp. 88–107). Berlin: Springer.

Shipp, T., Qi, Y., Huntley, R., & Hollien, H. (1992). Acoustic and temporal correlates of perceived age. *Journal of Voice, 6*(3), 211-216.

Voigt, R., Jurafsky, D., & Sumner, M. (2016). Between- and Within-Speaker Effects of Bilingualism on F0 Variation. In *Proceedings of Interspeech*, 1122-1126.

Walker, A. (2007). The effect of phonetic detail on perceived speaker age and social class. In J. Trouvain & W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS).*

Waller, S. S., Eriksson, M., & Sörqvist, P. (2015). Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in Psychology, 6*, 1-11. doi:10.3389/fpsyg.2015.00978

Weinberger, S. (2015). *Speech Accent Archive.* George Mason University. Retrieved from http://accent.gmu.edu

Yamazawa, H., & Hollien, H. (1992). Speaking fundamental frequency patterns of Japanese women. *Phonetica, 49*(2), 128-140.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. In (Eds.) Campbell, Gibbon, & Hirst, *Speech Prosody, 7*, 1037-1041.