Elderly listeners’ identification of Japanese long vowel pair ‘obasan’ and ‘obaasan’ using pitch and duration

C. T. Justine Hui* and Takayuki Arai
Graduate School of Science and Technology, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan

(Received 18 April 2018, Accepted for publication 18 October 2018)

Abstract: Japanese has long and short vowel distinction. While duration is the primary cue for listeners, pitch is being used as the secondary cue when duration becomes ambiguous. Duration however is affected by phonetic environment and therefore pitch cues may be more important in daily conversations. At the same time, ageing is known to affect speech recognition, in particular, pitch contour discrimination, such as tones. The current study compared a group of 15 young listeners with 14 elderly listeners using the words ‘obasan (aunt)’ and ‘obaasan (grandmother),’ manipulated in six steps duration-wise and pitch-wise. We found elderly listeners to use pitch more than the young listeners at the duration extremes, suggesting a generational effect on acceptability in accents (or lack of). At the same time, we observed half the elderly listeners to be less sensitive to pitch when duration becomes unreliable, depending on their fundamental frequency difference limens. The more sensitive elderly listeners, who performed similarly to the younger participants, significantly differed in their perception results from the less sensitive elderly listeners. This suggests that pitch deficits are present in half of the near-normal hearing elderly group, contributing to their inability to use pitch cues as well as their younger counterpart.

Keywords: Japanese long vowels, Pitch perception, Speech perception, Ageing

PACS number: 43.71.Lz, 43.71.Es, 43.66.Hg [doi:10.1250/ast.40.105]

1. INTRODUCTION

Japanese distinguishes between long and short vowels, where duration is an obvious auditory cue to distinguish between the two. In addition to duration, previous studies have concluded that pitch contour is used by native speakers to identify vowel quantities e.g. [1,2]. By using nonsense words with Japanese pitch patterns and a duration continuum, Kinoshita et al. [1] established that native speakers use duration as their primary cue to differentiate between vowel lengths, but pitch becomes important when the duration is ambiguous. The importance of teaching pitch contour to learners for perceiving Japanese long vowels has also been discussed by Deguchi [3], where he noted other studies have found duration cues to be less reliable with different speech rate [4], and depends on surrounding phonetic factors such as syllable positions [5].

Japanese learners with no vowel length distinctions and no tonal differences in their native language, such as English, tend to have more difficulties [2,6] in obtaining a native-like perceptions on vowel quantities. While Takiguchi [2] found that Chinese participants, where pitch is important in their native language, can perform almost as well as native Japanese listeners, Tsukada [7] did not find Arabic speakers, where vowel length is phonemic like Japanese, to be able to distinguish Japanese vowel length pairs.

These studies illustrate the importance of perceiving pitch cues in differentiating vowel length in daily conversations. On the other hand, most studies on Japanese vowel length distinction focus on the effect of native languages and nativity of Japanese, but little has been done on investigating the effect of ageing on vowel quantities perception.

It is well known that pitch perception is affected by age and hearing loss. Many previous studies have shown that frequency discrimination degrades with age. To name a few, Clinard has found significant declines in frequency difference limens (FDLs) at 500 Hz and 1,000 Hz as age increased [8], and Moore reported elderly normal hearing participants had very large FDLs for both pure and complex tones, despite near-normal auditory filters [9].

*e-mail: justinehui@eagle.sophia.ac.jp
In addition, ageing also makes speech recognition more difficult [10,11], possibly due to older listeners’ reduced spectral and temporal resolution [12,13]. In terms of the use of pitch in languages, Wang has found that ageing reduces the ability to categorically perceive pitch contour in Chinese [14], where elderly listeners displayed shallower slopes than younger listeners in tone identification.

We therefore hypothesised that elderly listeners will find it more difficult than their younger counterparts to differentiate between the two words, especially when duration becomes ambiguous and the use of pitch is needed. In the current study, we examined the differences between how young and elderly listeners utilise pitch and duration cues when identifying vowel lengths using one Japanese word pair ‘obaasan’ (grandmother) and ‘obasan’ (aunt).

2. METHODOLOGY

The minimal pair ‘obaasan (grandmother)’ and ‘obasan (aunt)’ were chosen as stimuli for this study as they differ in the length and pitch of /a/ and both words have similar familiarity for Japanese listeners. Real words were chosen to eliminate bias towards duration that may exist over pitch cues when asking participants to decide between a long and short vowel for a nonsense word.

Recordings were made by a female native speaker (Age: 27) from the Tokyo region speaking standard Japanese in a sound treated room with a sound recorder (Marantz PMD 671). The following sentences were recorded: ‘are ha obasan/obaasan desu’ (It is [my] grandmother/aunt over there.). The minimal pair was put into a carrier sentence to avoid effects of speech rate, allowing listeners to gauge the relative duration for long and short vowels. The experiment has been approved by the ethics committee at Sophia University.

Two additional tests were carried out for twelve of the elderly participants. They were the fundamental frequency (F0) difference limens test and the trail making test to check their cognitive ability.

2.1. Stimuli Manipulation

The words ‘obasan/obaasan’ within the sentence pair were manipulated using praat [15] where the duration, pitch or both were modified. Referring to Fig. 1, we have the original obasan/obaasan at the two diagonal ends of the grid. The ‘a’ in the bracket indicates that the stimulus was manipulated from the ‘obasan (aunt)’ original sound file, and the ‘g’ from the ‘obaasan (grandmother)’ original file.

Horizontally, we have the duration manipulation, where the length of /a/ is increased in 5 steps to become the same duration in the /a/ in ‘obaasan’ and vice versa. The /a/ in ‘obasan’ was measured to be 100 ms and the /a/ in ‘obaasan’ was 250 ms. Therefore each step increased or decreased by 30 ms. The ‘To Manipulation’ function in praat was used to create a Manipulation object, which is then used to modify the duration by adding duration points in the durationTier. The sound files were created by the ‘Publish resynthesis’ command.

Vertically, the pitch contour is modified using the PitchTier in the ManipulationEditor on the Manipulation object. The correct pitch accent for ‘obasan’ is LHHH and LHLLL for ‘obaasan,’ which can be observed from the original files, where the pitch contour lies flat between ‘oba-‘ and ‘-san,’ while in ‘obaasan,’ there is a decreasing slope in pitch at ‘-asan.’ The frequency values for the utterances are indicated in Fig. 2 where the top and the bottom line represents ‘obasan’ and ‘obaasan’ respectively.

Again, five step continual stimuli were created to match the pitch contour of ‘obasan’ to ‘obaasan’ and vice versa. As cent is perceived linearly by the human ears, the unit of cent was used to create the pitch continuum instead of Hertz as the unit of step, to allow comparison between duration and pitch in a more linear way. A hundred cent is equivalent to a semitone on the musical scale. The pitch difference at ‘a’ and ‘san’ was measured in Hertz and converted to cent, where each step was calculated to be 6 cent and 84.6 cent respectively at ‘a’ and ‘san.’ The pitch unit in the ManipulationEditor was first changed to semitones before the pitch points were added. The schematic diagram in Fig. 2 illustrates the pitch contour manipulation between ‘obasan’ and ‘obaasan.’ The pitch
of the original files were normalised prior to the manipulation to restrict variations in the sound files. To make the sentence sounds more natural, the ‘desu’ was modified accordingly. The audio files were also normalised in terms of their intensity to avoid effect of loudness on the perception of the words using the ‘Scale intensity’ function in praat.

2.2. Test Procedure

The young group carried out the test in the form of a Qualtrics web survey in accordance with a previous test to compare the perception between native and non-native listeners [16]. A web survey was chosen to accommodate non-native listeners residing outside of Japan for that study, irrespective of the current study. However, the elderly listeners had trouble manoeuvring the web survey and therefore the test was migrated onto PsychoPy [17] for its intuitive interface. Both groups carried out the test in a sound-treated room and listened to the stimuli over headphones (Sennheiser, HDA200) via a digital audio interface (Roland, Edirol UA-25EX).

The participants listened to the 36 tokens as specified previously with 6 repetitions in random order and answered the following question: ‘areha dare desu ka (Who is it over there?),’ by choosing either ‘Grandmother’ or ‘Aunt.’ The answers were provided in English, and for the elderly speakers, the descriptions ‘mother of parent’ and ‘sister of parent’ were included in Japanese, to focus on the semantics of the word and to avoid effect from the orthographic form in Japanese, where grandmother has an extra ‘a’ and aunt does not.

When the elderly listeners were unclear regarding the instructions, they were able to practise with two sample stimuli prior to the test. The stimuli were played back to the elderly listeners at initially 20 dB above their individual average threshold (500 Hz, 1,000 Hz, 2,000 Hz and 4,000 Hz), but since many claim to not hear it clearly, it was raised to 30 dB. The young listeners were allowed to adjust the signal level to a comfortable volume.

Two additional tests were carried out on twelve of the elderly participants. The F0 difference limens (F0DL) test gives us the just-noticeable-difference (JND) of fundamental frequency. The elderly participants were presented with two tones, a reference tone with a variable tone, where the variable tone changes according to the two-down, one-up adaptive procedure [18], using parts of the code from Soranza and Grassi’s Psychoacoustics Matlab package [19]. The relative FDL (the relative distance in frequency from the reference tone to the variable frequency tone) is obtained from taking the geometric means for 8 out of 12 reversals at three frequencies: 264.6 Hz and 207 Hz, which are the two frequency points where the ‘san’ changes from ‘obasan’ to ‘obaasan’ as shown in Fig. 2, as well as 392 Hz (G4 on the musical scale), which is roughly 1.5 times more than 264 Hz.

The second additional test examines the cognitive performance of the same twelve elderly participants as the F0DL test in the form of a trail making test [20]. The participants were given a screen of 25 digits in part A of the test and were asked to touch the screen in order from 1 to 25. The participants completed part A twice (with different location of the digits) before going onto part B, where there are now both digits and kana (Japanese syllabic writing system) from 1 to 13 and ‘a’ to ‘shi.’ The task required the participant to touch one digit followed by one kana in alternate order (1 - ‘a’ - 2 - ‘i’ and so on). Again, part B was completed twice with different locations of the digits and kana for each test. We gave all of the participants the same tests and recorded the time it took to complete each of the tests.

The trail making test (TMT) was chosen as it has been shown that TMT measures executive functions such as task switching, which predicts relative difficulty in understanding accented speech in older adults [21] and young adults [22], as well as speech in noise recognition [23]. A TMT programme was written specifically to be administered on a touch-screen for the participants to carry out the test and to measure the time to complete the test more meticulously.

2.3. Participants

Fifteen young listeners (mean age: 22.2) and fourteen elderly listeners (mean age: 68.9, three-frequency pure-tone audiometry (PTA): 15.5 dB HL) participated in this study. They were all native Japanese speakers. Half of the elderly listeners were not born in the Tokyo region, but all of them have been living in Tokyo for at least 30 years. Two of the participants from the young group were born outside of the Tokyo region in Hokkaido, one has been living in Tokyo for ten or more years, the other has lived within the Tokyo region for at least 5 years.

Auditory thresholds and filters were only measured for the elderly group and their results are shown in Figs. 3 and 4 and their values displayed in Table 1. The auditory filters in equivalent rectangular bandwidth (ERB) were taken using the Rion HD-AF as documented in [24] at 500 Hz, 1,000 Hz and 2,000 Hz. The level of the signal probe is set to be 20 dB above the threshold levels. Auditory thresholds were taken at 250 Hz, 500 Hz, 1,000 Hz, 2,000 Hz and 4,000 Hz. Figure 3 shows that the participants have mild high frequency sensorineural loss over all, and Fig. 4 displays the relative ERB with regards to the centre frequency (ERB/centre).

In addition, we obtained the F0 difference limens from twelve of the fourteen elderly participants, as well as their cognitive performances via trail making tests (ID: 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13).
The mean relative F0 difference limens of the three frequencies in percentage are shown in Fig. 5. The colours indicate whether or not the participant scored over the overall mean (18.83%) of the group. Seven participants showed roughly 10% and below relative F0DL, and the other participants roughly 20% and above.

In terms of cognitive performance, the mean time for part A of the trail making test was 54.98 s (SD = 22.7 s) and part B was 107.24 s (SD = 46.1 s). As a comparison, Hashimoto et al. have reported a mean time of 57.3 s (SD = 16.2) for part A and 154.4 s (SD = 71.7) for part B for participants age 70 to 74 with more than 10 years of education [25]. There was no significant difference ($t(5.8) = -0.8, p = 0.45$) between the two groups of elderly listeners as divided in Fig. 5.

### 3. RESULTS

The data is analysed mostly using logistic regression in the form of generalised linear mixed models using the R package *lme4* and function *glmer* [26]. Classification trees were also used to look at the importance of factors with the R package *rpart* [27] and plotting package *rpart.plot* [28] were used. Other packages used include *effect* [29] and *sjPlot* [30].

#### 3.1. Young Listeners

The young listeners group acted as the control group to compare the elderly listeners with. Figure 6 shows the average ratio of the young participants answering 'grand-

---

**Table 1** Auditory thresholds and ERB measurements of participants in the elderly group.

| ID | Age | Sex | Left ear | Right ear | Left ear | Right ear |
|----|-----|-----|----------|-----------|----------|-----------|
| 1  | 72  | F   | 9        | 25        | 5 23     | 25        |
| 2  | 68  | M   | 6.5 47   | 3 27      | 6.5 3 3.5 | 18 9      |
| 3  | 63  | M   | -1 -1.5  | -1 -1.5   | -3 -3.5  | -3 3      |
| 4  | 60  | F   | 8.5 13   | 8.5 13    | 8.5 13   | 13 13     |
| 5  | 71  | M   | 0.5 7    | 0.5 7     | 2.5 5    | 2.5 5     |
| 6  | 72  | F   | 10.5 25  | 24.5 35   | 14.5 27  | 14.5 27   |
| 7  | 67  | F   | 24.5 41  | 20.5 31   | 14.5 29  | 14.5 29   |
| 8  | 75  | F   | 32.5 43  | 22.5 31   | 14.5 29  | 14.5 29   |
| 9  | 80  | F   | 22.5 25  | 28.5 31   | 14.5 29  | 14.5 29   |
| 10 | 70  | F   | 2.5 11   | 8.5 11    | 14.5 29  | 14.5 29   |
| 11 | 69  | F   | 6.5 19   | 12.5 17   | 14.5 29  | 14.5 29   |
| 12 | 63  | M   | 0.5 7    | 10.5 7    | 14.5 29  | 14.5 29   |
| 13 | 73  | M   | 12.5 15  | 18.5 19   | 14.5 29  | 14.5 29   |
| 14 | 62  | F   | 17 31    | 22.5 29   | 14.5 29  | 14.5 29   |

**Fig. 3** Hearing thresholds of elderly listeners.

**Fig. 4** Auditory filters in the form of relative equivalent rectangular bandwidth (ERB) at 20 dB above the participant’s threshold levels.
mother’ for each stimulus. The opaqueness of the tiles reflects the ratio of the responses to be ‘grandmother.’

Using a generalised linear mixed model (GLMM) with a logit link, we obtained the following odd ratios by exponentiating the coefficients of the model, where the fixed effects were duration and pitch, and the random effect was the participant in Table 2.

The odd ratios from Table 2 shows that one unit increase in duration is more than six times likely to increase the odds of perceiving the stimulus as ‘grandmother’ than one unit increase in pitch. In terms of interaction between pitch and duration, we did not find any significant effect for young listeners ($\chi^2(1) = 2.56, p = 0.11$).

### Table 2  Odd ratios of young group from the GLMM.

|                | Intercept | pitch | duration | pitch:dur |
|----------------|-----------|-------|----------|-----------|
| Estimates      | 0         | 2.79  | 12.18    | 1.07      |
| 95% CI         | 0,0       | 2.10-3.74 | 1.22,8.73 | 0.98,1.17 |

Figure 7 shows the predicted probabilities obtained from the GLMM for the young group using the R package `effect` [29]. Along the $x$-axis we have the pitch levels, and each line represents a different duration level. $y$-axis is the predicted probabilities of the answer being ‘grandmother’ (i.e., 1 being ‘grandmother,’ and 0 being ‘aunt’). The shaded areas indicate the 95% prediction bands.

As we can see from Fig. 7, the young listeners do not seem to use pitch cues as much as duration cues, supporting the evidence from the odd ratios in Table 2. For example, at the pitch level of ‘obasan’ (P0), the D5 (grandmother) line shows that most participants regarded the stimulus as ‘grandmother’ despite the pitch accent being the same as ‘aunt.’ The same can be seen on the other extreme end at P5, where the answer of the young listeners are governed by the duration cues.

While the extremes of duration levels are not affected by pitch cues, once the duration becomes ambiguous at D2, pitch cues appear to be more important for the young listeners. The D2 line exhibits the sharp transition from ‘aunt (0)’ to ‘grandmother (1)’ where the listeners resort to using pitch as the auditory cue for deciding between the answers.

### 3.2. Elderly Listeners

Figure 8 shows the average ratio of the elderly participants answering ‘grandmother’ for each stimulus.

---

**Fig. 5** Relative F0 difference limens of 12 elderly listeners.

**Fig. 6** Average ratio of answer as ‘grandmother’ for young listeners.

**Fig. 7** Predicted probabilities from the GLMM for young group in terms of pitch and duration.
The opaqueness of the tiles reflects the ratio of the responses to be ‘grandmother.’

Similar to the young group, we used a generalised linear mixed model (GLMM) with a logit link, where duration and pitch were fixed effects and ID of the participants was the random effect. We found a significant interaction effect between duration and pitch for the elderly listeners (\( \chi^2(1) = 12.2, p = 0.00048 \)).

The odd ratios, from exponentiating the coefficients returned by the GLMM are shown in Table 3. From the odd ratios, we can see that one unit increase in duration is roughly twice as likely to increase the odds of perceiving the stimulus as ‘grandmother’ than pitch.

Figure 9 shows the predicted probabilities from the GLMM for the elderly group, where the probabilities are generated using the R package \textit{effects} [29].

Visually inspecting Fig. 9, we can see that unlike the young group, even at the longest duration at D5 (obaasan/grandmother), some participants still used the pitch cues and regarded the stimulus as ‘aunt.’ This suggests that the elderly speakers, at least at the extreme end of the pitch levels at P0 (obaasan/aunt), used pitch cues more than their younger counterparts. The effect of pitch cues are seen less prominently on the other side of the extreme at P5 (obaasan/grandmother). On the other hand, we do not see the steep change from ‘aunt’ to ‘grandmother’ as in the young group, for when duration is ambiguous at D2. This suggests that while the elderly groups consider pitch cues to be more important than the younger group at the extreme, they may also have more trouble using the pitch cues than their younger counterparts when duration is ambiguous.

### 3.2.1. Hearing thresholds of elderly group

The average threshold across all frequencies did not have an effect on how the elderly group performed. However, we found some significant effects from the auditory thresholds of specific frequencies. Unfortunately a model with all the threshold levels at each frequency, pitch and duration did not converge and therefore we analysed the effects of the thresholds and pitch at each level of duration separately. The models were chosen by comparing the Akaike information criterion (AIC) obtained from different combination of threshold averages at each frequency. The significant effects of auditory thresholds at specific frequencies are shown in Table 4.

Table 4 shows that at longer duration levels (towards ‘grandmother’), auditory thresholds at 500 Hz made a difference to what the participant perceives. The only frequency that made a difference at the shortest duration

---

**Table 3** Odd ratios of elderly group from the GLMM.

| Est   | Intercept | pitch | duration | pitch:duration |
|-------|-----------|-------|----------|----------------|
| 0.008 | 1.28      | 2.35  | 1.079    |
| 0.004,0.017 | 1.09,1.50 | 2.01,2.76 | 1.03,1.13 |

**Table 4** Frequency of auditory threshold at which there is a significant effect.

| duration level | Frequency of auditory threshold                                      |
|----------------|---------------------------------------------------------------------|
| 0              | 250 Hz \( \chi^2(1) = 8.12, p = 0.0044 \)                              |
| 3              | 500 Hz \( \chi^2(1) = 9.09, p = 0.0026 \)                               |
|                | 1,000 Hz \( \chi^2(1) = 4.31, p = 0.038 \)                             |
| 4              | 500 \( \chi^2(1) = 9.34, p = 0.0022 \)                                 |
|                | 2,000 Hz \( \chi^2(1) = 6.77, p = 0.0093 \)                            |
| 5              | 500 Hz \( \chi^2(1) = 21.6, p < 0.0001 \)                              |
|                | 2,000 Hz \( \chi^2(1) = 12.6, p < 0.0001 \)                           |
level (at D0) was 250 Hz. The results also show that higher frequencies (1,000 Hz and 2,000 Hz) are more important at longer duration levels. We could not converge the model to look at the auditory filters results, and therefore the differences in ERB will be discussed in Sect. 3.4.

3.3. Comparison in Terms of Age

Table 5 shows the odd ratios of the combined model of young and elderly listeners. We found significant interaction effects between pitch and group ($\chi^2(1) = 129.03, p < 0.0001$) as well as between duration and group ($\chi^2(1) = 288.32, p < 0.0001$), where group separates the young listeners from the elderly listeners.

Figure 10 created using R package sjPlot [30] demonstrates the interaction between the young and elderly group at each pitch level over the duration levels and Fig. 11 shows the interaction at each duration level over the pitch levels.

In Fig. 10, we can observe a shift of slope for the young group towards lower number in duration (towards ‘obasan’), as pitch level increases (towards ‘obaasan’), while the slope of the elderly group becomes steeper as the pitch level increases and does not resemble an ‘s-shaped’ in general. This shows that for the young group, change in pitch appears to go hand in hand with duration changes. As pitch increases, the younger group gave more ‘grandmother’ answers with longer duration. The same slope shape in all six pitch levels suggests that the change in perceiving ‘aunt’ to ‘grandmother’ is triggered mostly by duration. For the elderly group, while duration also affected the answers, the increase in pitch levels made it more likely for the group to perceive the stimulus as ‘grandmother.’ The change in slope shape implies that duration changes are different depending on the pitch level, supporting our analysis for the elderly group, where we found an interaction effect between duration and pitch.

Figure 11 shows that for each duration level over pitch changes, only D1 and D2 exhibit interaction between the two groups. To investigate further the differences between the young and elderly group, we concentrated on the D2 line, where duration is considered to be ambiguous as seen from the responses.

As previous studies have noted, when duration becomes ambiguous, native speakers rely on pitch cues to distinguish vowel length [1]. While technically speaking, D3 should also be considered to be ambiguous duration, looking at Fig. 7, we can see that it is D2 where young listeners make the categorical change from aunt to grandmother most clearly. Comparing the two slope at the 0.5 crossing point in Fig. 12, the young group has a slope of 0.41, and the elderly group has a slope of 0.18. This confirms our hypothesis that elderly listeners may be less sensitive to pitch cues than young listeners once the duration is deemed unreliable.

| Table 5 | Odd ratios from the combined (young and elderly) model [95% CI]. |
|---------|---------------------------------------------------------------|
| Intercept | pitch | duration |
| 0.026 [0.016, 0.044] | 1.38 [1.24, 1.55] | 2.56 [2.29, 2.88] |
| group | pitch:duration | pitch:group |
| 0.008 [0.003, 0.02] | 1.078 [1.04, 1.12] | 2.11 [1.83, 2.43] |
| duration:group |
| 4.87 [3.92, 6.14] |

**Fig. 10** Predicted probabilities in terms of duration comparing young group with elderly group.
3.3.1. Pitch or duration

We used a classification tree to look at how each group uses the pitch and duration cues differently. The classification tree looks for the variable that divides the responses into maximally different groups at each node until it can no longer divide further. This analysis follows the procedure as noted in [31].

The responses for each group of listeners were analysed with the \textit{rpart} function in R [27]. The dependent variable was the answer (aunt or grandmother) and the prediction factors were duration (D0 to D5), and pitch (P0 to P5). Since the young group did not have auditory threshold or ERB measured, these measurements were not included in the tree for the elderly listeners. The numbers in the top row of the bottom nodes represent the number of responses, left being the number of responses as ‘aunt,’ and the right ‘grandmother.’ The percentage is the percentage of observations in the dataset. \textit{rpart} uses K-fold cross validation and the complexity parameter with the smallest cross validation error was used to prune the trees as shown in Figs. 13 and 14.

Figures 13 and 14 display the classification trees for elderly and young group respectively. Both groups split initially at the same duration level (between 2 and 3). However, as we see pitch coming into play in the next level for the elderly group, it does not appear to be an important factor for the young group until the third and last split after duration has been accounted for. This shows that while the elderly group may have trouble perceiving pitch changes as well as the young group, pitch cues are more important to distinguish vowel length for the elderly group, thus supporting the analysis from the GLMM where there is interaction between pitch and duration for the elderly group.
group. Pitch only becomes important for the young group at D2, as duration is deemed too ambiguous to distinguish between ‘aunt’ and ‘grandmother.’

3.4. Effect of Fundamental Frequency Difference Limens (F0DL)

We carried out an additional experiment to look at the fundamental frequency difference limens (F0DL) from twelve of the fourteen elderly listeners. Their mean results in relative F0DL are shown in Fig. 5. Concentrating on how the two groups perceived the stimuli at D2 where duration is deemed ambiguous for the younger group, we split the elderly group into two sub-groups according to their F0DL results. They were considered ‘poor’ if their F0DL results were below the group mean. Their average results of identifying between ‘aunt’ and ‘grandmother’ for when duration is ambiguous are shown in Fig. 15.

Visually we can observe that the ‘good’ elderly group performs close to the young group, whereas the ‘poor’ group hovers around 50%, suggesting that they did not use pitch to decide between the word pair when duration becomes ambiguous. Statistically, while we found no significant difference between the ‘good’ elderly group and the young group ($t = 2.24$, $df = 4.7$, $p$-value = 0.078), there is a significant difference between the ‘good’ and ‘poor’ group in terms of their answers ($\chi^2(1) = 4.13$, $p = 0.04$).

Since the ‘good’ elderly group had no significant difference with the young group, we will focus on examining the other hearing properties within the elderly participants between the ‘good’ and ‘poor’ groups. Using the Welch’s $t$-test, we found no significant difference between the groups for their mean pure-tone audiometry for 500 Hz, 1,000 Hz and 2,000 Hz ($t = 2.24$, $df = 4.7$, $p$-value = 0.078). There was also no significant difference between their ERB values for 500 Hz ($t = 0.74545$, $df = 9.9922$, $p$-value = 0.4732), 1,000 Hz ($t = 2.0567$, $df = 6.1395$, $p$-value = 0.08438) and 2,000 Hz ($t = 0.49015$, $df = 9.9909$, $p$-value = 0.6346).

4. DISCUSSION

4.1. Effect of Hearing Properties

In the current study, we found that while there were elderly listeners who had trouble distinguishing between the word pair when duration becomes ambiguous, there were also elderly listeners who performed similarly to their younger counterpart. We found that while the two groups do not differ significantly in terms of their pure-tone audiometry or auditory filter bandwidths, the young-like elderly group performs significantly better than the poorer group in terms of their fundamental frequency difference limens.

It is also interesting to note that our ‘poor’ group, despite near-normal hearing thresholds, had a mean relative FDL of 35.93%, compared to 10% reported in [9], where hearing impaired elderly listeners were tested. It is possible that the ‘poor’ group of elderly listeners are originally amusics, meaning that they already lack the ability to distinguish between pitches. However past literature has found that the mean FDL for young amusics to be 5%, which is still below the relative FDLs from the current study, even for the ‘good’ group. This suggests that some elderly listeners may have greater deterioration in pitch discrimination, which affect their daily speech perception in realistic circumstances (e.g. fast speech with shortened long vowels).

4.2. Acceptableness of Incorrect Prosody

In Japanese, a long ‘obasan (aunt)’ with a LHHHH pitch accent and a short ‘obaasan (grandmother)’ with a LHLL pitch accent do not exist formally. However we saw
that almost all participants in the young group deemed ‘obaasan’ with a long /a/ regardless of pitch to be ‘grandmother,’ and ‘obasan’ with a short /a/ to be ‘aunt.’ On the other hand, the elderly listeners appear to take pitch more into consideration at least at the upper duration extreme (D5) with at least 25% of the answers regarded as ‘aunt’ when the pitch accent is LHHHH. This can also be confirmed in the classification tree in Fig. 13, where for higher degree of duration (closer to ‘obaasan’), pitch is the more deciding factor, as opposed to the tree for the younger group, where pitch only comes into play when duration is 2 (i.e. ambiguous).

One explanation for this may be a case of increasing acceptableness of incorrect or unaccented prosody in the younger generation since we also saw during ambiguous duration (D2), the younger generation are capable of using pitch as the cue to decide between ‘aunt’ and ‘grandmother.’ This is referred as ‘heiban-ka’ or deaccentualisation of nouns, a phenomenon observed in recent Tokyo dialect changes [32]. Lee and Takano have also reported seeing similar acceptance increasingly in younger speakers in foreign loan words and other daily life words [33,34].

Meanwhile, the elderly group is less tolerant of incorrect prosody and therefore we see that pitch is still being used as a cue together with duration at the upper duration extreme (D5). Comparing between the elderly groups 1 and 2, where the groups were split according to their use of pitch perception when duration becomes ambiguous at D2, we found no significant effect of the groups ($\chi^2(1) = 0.48, p = 0.49$) at D5. This suggests that while hearing deficit may be present between the two groups resulting in the differences in pitch perception at D2, it appears that the difference in behaviour we see at D5 between the young and the elderly listeners is more likely due to generational differences.

Previous research has also showed that speakers of syllabeme dialects in Japan have a tendency to perceive short and long syllable more gradually than mora dialect such as Tokyo dialect (standard Japanese). However, within our elderly population we only had one speaker from Akita, who moved to Tokyo at the age of 18 and therefore we did not analyse the data with regards to syllabeme dialect closely.

### 4.3. Stimuli Creation

The voice of stimuli may also have an effect, as we used a young female speaker, which is much closer in age than the young group to the elderly group. The voice may appeal more to the younger group and thus allowing them better performance during ambiguous duration. However, as there is no significant difference in age between group 1 and group 2, the effect of the age of the speaker may be minor in this case.

### 4.4. Longer Duration Clearer for Elderly Listeners

Previous studies have also suggested that longer duration may enhance speech intelligibility for elderly listeners, such as the effect of clear speech [35], and speech rate [36], where Gordon-Salant has shown that longer cues help elderly listeners to accurately discriminate and identify speech sounds. In terms of pitch contours, Wang [37] showed that longer duration helps older listeners to categorically perceive flat-rising tones in Mandarin. This could also be interpreted similarly in the current study as shown in Fig. 13, where the elderly responses were divided using pitch cues at higher duration levels, compared to lower duration levels. Separating the GLMM into two models with the first model consisting of the duration levels from 0 to 2 and the second model from 3 to 5, the odd ratios for pitch effect is larger at the longer durations (1.6) than the shorter durations (1.4). Further studies will be needed to confirm this observation of longer duration allowing elderly listeners to make use of the pitch cues more readily for vowel length distinction in Japanese.

### 5. CONCLUSIONS

The current study examined the difference between young and elderly listeners to categorically perceive the long-short vowel pair ‘obasan-obaasan (aunt-grandmother).’ Both groups appeared to use duration as their primary cue, and pitch cues as their secondary cue once duration was considered to be unreliable. Elderly listeners tend to use more pitch cues than young listeners at the duration extremes, suggesting differences in acceptability in accent due to generation gap. At the same time, roughly half of the elderly listeners were less sensitive to pitch changes once duration became ambiguous. The other half of the elderly listeners were able to perceive pitch changes as well as the young group, and we saw significant differences between these two groups in their responses when perception in pitch changes is most needed. While the two groups did not differ in hearing threshold or auditory filter bandwidth, they had significant differences in their frequency difference limens results. This shows that instead of generational differences, pitch discrimination deficits were to be accounted for the shallower slope displayed by half the elderly listeners. As we only examined one word pair with short-long vowel distinction, the results in this study cannot be generalised as the universal behaviour of long vowel identification in Japanese due to its limitation in accent type and position of the vowel. However, results from the current study gave evidence of pitch processing degradation affecting speech perception in less than optimal situations, and we may be able to apply the current findings to future work in a wider context.
ACKNOWLEDGEMENT

We would like to thank our anonymous reviewers and the editor for their helpful comments, Douglas Hing for providing us with the Trail Making Test, our speaker for the recording, and our participants for their patience and time.

REFERENCES

[1] K. Kinoshita, D. M. Behne and T. Arai, “Duration and F0 as perceptual cues to Japanese vowel quantity,” Proc. 7th Int. Conf. Spoken Language Processing, (1), pp. 757–760 (2002).
[2] I. Takiguchi, “Perceptual development on the identification of length in L2 Japanese,” 17th Int. Congr. Phonetic Sciences (ICP18 XVII), pp. 1950–1953 (2011).
[3] M. Deguchi, “The role of pitch contours in teaching vowel length distinctions in Japanese,” Proc. 6th Pronunciation in Second Language Learning and Teaching Conf., pp. 174–182 (2015).
[4] Y. Hirata, “Effects of speaking rate on the vowel length distinction in Japanese,” J. Phone., 32, 565–589 (2004).
[5] Y. Minagawa, T. Kagomiya and K. Maekawa, “Durational variations of Japanese long/short vowels in different speaking rates: Analysis of a spontaneous speech corpus,” ICPkS, pp. 2127–2130 (2003).
[6] K. Tajima, H. Kato, A. Rothwell, R. Akahane-Yamada and K. G. Munhall, “Training English listeners to perceive phonemic length contrasts in Japanese,” J. Acoust. Soc. Am., 123, 397–413 (2008).
[7] K. Tsumada, “Non-native Japanese listeners’ perception of vowel length contrasts in Japanese and Modern Standard Arabic (MSA),” Second Lang. Res., 28, 151–168 (2012).
[8] C. Clinard, K. L. Tremblay and A. Krishnan, “Aging alters the perception and physiological representation of frequency: Evidence from human FFR recordings,” Hear. Res., 264, 48–55 (2010).
[9] B. C. Moore and R. W. Peters, “Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity,” J. Acoust. Soc. Am., 91, 2881–2893 (1992).
[10] K. H. Arehart, P. E. Souza, R. K. Muralimanohar and C. W. Miller, “Effects of age on concurrent vowel perception in acoustic and simulated electroacoustic hearing,” J. Speech Lang. Hear. Res., 54, 190–210 (2011).
[11] D. Fogerty, D. Kewley-Port and L. E. Humes, “The relative importance of consonant and vowel segments to the recognition of words and sentences: Effects of age and hearing loss,” J. Acoust. Soc. Am., 132, 1667–1678 (2012).
[12] P. J. Fitzgibbons and S. Gordon-Salant, “Age effects on duration discrimination with simple and complex stimuli,” J. Acoust. Soc. Am., 98, 3140–3145 (1995).
[13] S. Gordon-Salant, G. H. Yeni-Komshian, P. J. Fitzgibbons and J. Barrett, “Age-related differences in identification and discrimination of temporal cues in speech segments,” J. Acoust. Soc. Am., 119, 2455–2466 (2006).
[14] Y. Wang, X. Yang, H. Zhang, L. Xu, C. Xu and C. Liu, “Aging effect on categorical perception of Mandarin tones 2 and 3 and thresholds of pitch contour discrimination,” Am. J. Audiol., 26, 18–26 (2017).
[15] P. Boersma and D. Weenink, “Praat: Doing phonetics by computer,” Available at http://www.fon.hum.uva.nl/praat/ (accessed 28 Jan. 2019).
[16] C. T. J. Hui and T. Arai, “Pitch and duration as auditory cues to identify Japanese long vowels for Japanese learners,” Proc. Autumn Meet. Acoust. Soc. Jpn., pp. 1465–1468 (2017).
[17] J. Peirce, “PsychoPy - Psychophysics software in Python,” J. Neurosci. Methods, 162(1-2), 8–13 (2007).
[18] H. Levitt, “Transformed up-down methods in psychoacoustics,” J. Acoust. Soc. Am., 49, 467–477 (1971).
[19] A. Soranzo and M. Grassi, “Psychoacoustics: A comprehensive MATLAB toolbox for auditory testing,” Front. Psychol., 5, 1–13 (2014).
[20] R. M. Reitan, “The relation of the trail making test to organic brain damage,” J. Consult. Psychol., 19, 393–394 (1955).
[21] P. Adank and E. Janse, “Comprehension of a novel accent by young and older listeners,” Psychol. Aging, 25, 736–740 (2010).
[22] L. Tao and M. Taft, “Influences of cognitive processing capacities on speech perception in young adults,” Front. Psychol., 8 (FEB) (2017).
[23] R. J. Ellis, P. Molander, J. Rönnberg, B. Lyxell, G. Andersson and T. Lunner, “Predicting speech-in-noise recognition from performance on the trail making test,” Ear Hear., 37, 73–79 (2016).
[24] T. Nakaichi, K. Watanuki and S. Sakamoto, “A simplified measurement method of auditory filters for hearing-impaired listeners,” Acoust. Sci. & Tech., 24, 365–375 (2003).
[25] R. Hashimoto, K. Meguro, E. Lee, M. Kasai, H. Ishii and S. Yamaguchi, “Effect of age and education on the Trail Making Test and determination of normative data for Japanese elderly people: The Tajiri Project,” Psychiatry Clin. Neurosci., 60, 422–428 (2006).
[26] D. Bates, M. Mächler, B. M. Bolker and S. C. Walker, “Fitting linear mixed-effects models using lme4,” J. Stat. Softw., 67, 1–48 (2015).
[27] T. Therneau and B. Atkinson, ?rpart: Recursive partitioning and regression trees,” Available at https://cran.r-project.org/package=rpart (accessed 28 Jan. 2019).
[28] S. Milborrow, ?rpart.plot: Plot ?rpart? models: An enhanced version of ?plot.rpart?,” Available at https://cran.r-project.org/package=rpart.plot (accessed 28 Jan. 2019).
[29] J. Fox, “Eect displays in R for generalised linear models,” J. Stat. Softw., 8, 1–27 (2003).
[30] D. Lüdecke, ?sjPlot: Data visualization for statistics in social science,” Available at https://cran.r-project.org/package=sjPlot (accessed 28 Jan. 2019).
[31] M. Brunelle, “Tone perception in Northern and Southern Vietnamese,” J. Phon., 37, 79–96 (2009).
[32] K. Maekawa, “Linguistic studies on the variation of speech (in Japanese),” Acoust. Soc. Jpn. (J), 52, 552–557 (1996).
[33] H.-R. Lee, “Pattern investigation of 65-year Period change of Mandarin Chinese tones 12 and tones 14: Effects of aging and hearing impairment,” 9–24 (2016).
[34] S. Gordon-Salant, “Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing,” J. Acoust. Soc. Am., 80, 1599 (1986).
[35] S. Gordon-Salant, “Hearing loss and aging: New research findings and clinical implications,” J. Rehabil. Res. Dev., 42, 9–24 (2006).
[36] Y. Wang, X. Yang and C. Liu, “Categorical perception of Mandarin Chinese tones 12 and tones 14: Effects of aging and signal duration,” J. Speech Lang. Hear. Res., 60, 3667–3677 (2017).