EFFECT OF FREQUENCY DISCRIMINATION ABILITY OF ELDERLY LISTENERS ON JAPANESE PITCH ACCENT PAIRS

C. T. Justine Hui[1], Yukiko Sugiyama[2], Takayuki Arai[3]

Sophia University[1,3], Keio University[2]
justinehui@eagle.sophia.ac.jp[1], yukiko@keio.jp[2], arai@sopahia.jp[3]

ABSTRACT

Hearing typically changes with age, affecting how listeners perceive speech, in particular, subtle cues that exist in languages. Japanese is a pitch-accent language with minimal pairs realised by their pitch contours, requiring listeners to pick up on such cues in pitch primarily. In the current study, we measured a group of near-normal hearing Tokyo born elderly listeners for their difference limens for frequency (DLF) and their ability to identify two pitch accent pairs *hana* "flower/nose" and *sita* "down/tongue" with 10-step continua. We found elderly listeners with better DLF to distinguish between the pairs, producing clearer results for the *hana* pair. The elderly listeners with poorer DLF were not able to separate between the accented and unaccented stimuli for either pair. This suggests that despite their similar near-normal hearing thresholds, elderly listeners differ in how they perceive pitch accent pairs depending on their ability to discriminate pure tones.

Keywords: pitch accent, speech perception, ageing, difference limens, hearing impairment

1. INTRODUCTION

As hearing declines with age, it often leads to difficulty in understanding speech. Nonetheless, previous studies have found that as long as the speech is audible and familiar, elderly listeners exhibit no particular deficits in speech intelligibility [34]. Difficulty arises when elderly listeners have to listen in adverse conditions, such as lower audibility, distorted speech signal including reverberant speech [16, 24], background noise [9, 15], faster speech rate [33], and talker's foreign accent [1]. While the degradation of the speech signal compromises speech intelligibility for elderly listeners, suprathreshold deficits in temporal [11, 13] and spectral processing [21] have also been found in the ageing population, affecting their ability to comprehend speech. Moore and Peters found generally broadened auditory filters and reduced abilities to discriminate frequency in their normal hearing elderly listeners [21]. Moore has also found difference limens to increase with age [20]. These findings can be observed in Mandarin tone recognition, where elderly listeners were reported to have shallower slope in identifying the tones, smaller peaks in the discrimination task, and higher threshold for pitch contour discrimination [30].

While Japanese does not have tones like Mandarin, it is a pitch-accent language, where pitch accent is realised by a sharp fundamental frequency (f0) drop in the accented word to the following particle, with the obvious primary cue being pitch [4, 28, 18, 29]. Previous studies have tried to examine whether or not secondary cues such as duration and intensity exist for Japanese pitch accent, with no conclusive agreement. Literature in general agrees that duration does not correlate with accent, but some studies have found a weak link to intensity [4, 7, 31]. Using minimal pairs of pitch accent without f0 information, Sugiyama [29] found listeners were able to differentiate between the minimal pairs at a rate better than chance. With evidence of elderly Mandarin listeners finding differences in tones to be difficult to perceive, elderly Japanese listeners may also have trouble deciphering pitch cues, especially if they have deficits in frequency discrimination. The current study has two objectives: 1. to examine how elderly listeners perceive pitch accent and how it relates to their frequency discrimination abilities; and 2. to find out whether listeners with poorer pitch processing abilities are able to differentiate between accented and unaccented words using potential secondary cues.

2. METHODOLOGY

The participants were first tested for their hearing abilities such as auditory threshold, auditory filter bandwidths and difference limens in pure tones. This was followed by an identification test of two sets of word pairs. Each word pair consists of two 10-step continua created from the accented original word and the unaccented original word, with only their pitch manipulated. The purpose of testing stimuli created from both original words was to investigate our second objective, whether or not secondary cues, such as duration or intensity, were present and perceivable, especially when some of the listeners have poorer pitch processing abilities.

2.1. Participants

Twenty-one elderly Tokyo-born Japanese listeners (female: 11) participated in the current study. Their age ranged from 60 to 83 years old (*M* =72.1, *SD*=6.1) and their pure tone audiometry (PTA: mean of 500 Hz, 1k Hz and 2k Hz) was 15 dB HL (*SD*=7.7). The participants were then separated into two groups according to whether their DLFs results were higher or lower than the median (0.73%) as shown in Fig. 1. The DLFs of the two groups were significantly different ($t(9.5)=-3.8$, $p<0.01$). The two groups had a median of 0.58% and 1.25% respectively. As a reference, normal hearing subjects were reported to have a mean DLF of 0.75% [23] and 1.2% for hearing impaired elderly listeners [22].

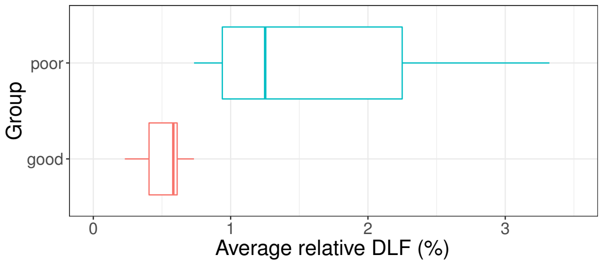

Figure 1: DLFs of the two groups

2.2. Stimuli Creation

The two sets of word pairs were *hana* (LH) meaning flower when accented and nose when unaccented, and *sita* (LH) meaning tongue when accented and down when unaccented. Both word pairs have a similar familiarity index according to [2, 29]: flower (6.5) versus nose (6.4) and tongue (6.1) versus down (5.9) on a 7-point scale with 7 being most familiar. The original files were recorded by a female Tokyo-born speaker. Four sets of 11 continual stimuli (10 linear steps measured in cents) were created from four original files by manipulating the pitch from the accented words to unaccented and vice versa. The carrier phrase, *watashi wa* [target] *ga suki* "I like [target]", was kept the same per word pair. When the target word is followed by the phrase *ga suki*, the *suki* is subject to catathesis, the compression of f0 range, if the preceding word within the same accentual phrase is accented [28]. The effect of catathesis could be observed in the original accented phrases. In the current study, to keep the carrier phrase constant within word pairs, only the carrier sentences from the original unaccented stimuli were used for both *hana* and *sita*. Pitch manipulations were carried out with Praat [5] using the 'To Manipulation' function to create a Manipulation object and the pitch contour was modified using the 'PitchTier' in the 'ManipulationEditor'. The sound files were synthesised by the 'Publish resynthesis' command. Altogether, there were 22 stimuli per word pair and the resulting stimuli were visually and aurally checked by the second author, a phonetically trained native Japanese listener, to make sure that they sound natural and were easily differentiable.

2.3. Test procedure

The participants first completed an auditory threshold assessment for both ears at 250 Hz, 500 Hz, 1k Hz, 2k Hz, and 4k Hz. Only the better ear was used from then on. They were measured for their auditory filter bandwidths at 500 Hz, 1k Hz, and 2k Hz using the Rion HD-AF [25]. Finally, they carried out the frequency discrimination test to measure their difference limens (DLFs) at 250 Hz, 500 Hz, 1k Hz and 2k Hz followed by an identification test for the pitch accent pairs. The frequency discrimination test stimuli were presented at 30 dB SL for each frequency, and the speech stimuli were presented at 30 dB above their PTA at 500 Hz, 1k Hz and 2k Hz. Their cognitive level was confirmed in the form of a trail making test (TMT), which was reported to predict speech processing for elderly listeners [10, 12]. The participants carried out the test in a sound-treated room, where they listened to the stimuli over headphones (Sennheiser, HDA200) via a digital audio interface (Roland, Edirol UA-25E) and responded by using a touch screen.

2.3.1. *Difference limens for pure tones (DLFs)*

The frequency discrimination threshold for pure tones (DLFs) were measured using a two-interval two-alternative forced-choice method, as described in [23, 22]. Two intervals were played to the participant, one with four successive tones A at a fixed frequency (AAAA), the other interval consisted of tone A and B played alternately (ABAB). Both tone A and B were 300-ms tones with 20-ms raised-cosine ramps, with a 100 ms silent interval between the two intervals. Tone B had a higher frequency than tone A, where the difference in the frequencies of the two tones were adjusted adaptively in a two-down-one-up procedure. The order of the intervals was randomised at each trial, and the participant was asked to judge whether the ABAB interval came first or second by selecting "1" or "2" on the screen. The procedure ran for eight reversals and the listeners' thresholds were estimated as the geometric mean of the step for the last six reversals. For participants who performed more than 10 trials correctly in succession within the last six reversals, only the remaining reversals were considered. The participants were made sure to understand the task by completing a practice session of 10 trials or more if needed. Feedback was given for the practice session.

2.3.2. Pitch Accent

The participants listened to 22 stimuli in random order per word pair to decide whether they heard the accented or the unaccented stimulus. Each block was repeated five times. They responded by choosing one of two alternatives in *kanji* (Chinese characters) on the screen. Half of the listeners started with the *hana* pair and half with *sita*. Before the task, they were given a practice session with five unprocessed stimuli of *hati* "eight/bee".

3. RESULTS

The mean ratios of answers as the accented word ("flower" for *hana*, and "tongue" for *sita*) for the 'good' group are illustrated in Fig. 2 and the 'poor' group in Fig. 3. The 'good' group was able to distinguish "flower" from "nose" for the *hana* pair relatively well with a steep s-curve. While the listeners in this group could not separate the two *sita*'s as well, we can still observe an s-curve-like shape in their responses. On the other hand, the 'poor' group had a much less coherent pattern in both *hana* and *sita* identification, showing that they had trouble identifying the accented and the unaccented word. The different point types show the two types of original files used as the stimuli in the figures.

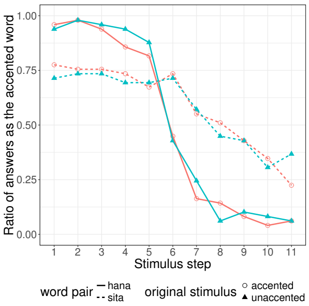

Figure 2: Good DLF

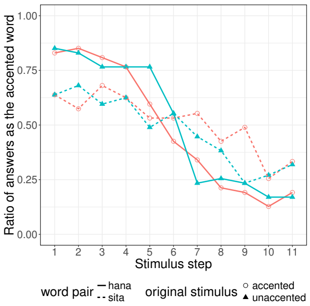

Figure 3: Poor DLF

While the two groups differed significantly in their DLFs, there were no significant differences between the two groups in terms of PTA (average of 500 Hz, 1k Hz and 2k Hz) ($t(18) = 1.4$, $p = 0.19$), the reaction time from the trail making test ($t(16.3) = -1.0$, $p = 0.36$), or their auditory filter bandwidths in equivalent rectangular bandwith (ERB) for 500 Hz ($t(12) = 1.3$,$p = 0.23$), 1k Hz ($t(17) = 0.11$, $p = 0.9$) or 2k Hz ($t(11) = -0.74$, $p = 0.48$).

3.1. Effect of DLF results on pitch accent perception

We analysed the results with logistic mixed models using the *glmer* function in the R package *lme4* [3]. Likelihood ratio tests were carried out to examine effects of the DLF groups, word pair, stimulus step and whether or not the stimulus was modified from an unaccented or accented original stimulus on the response (whether the participant perceived an accented or unaccented word). As the original stimulus factor causes the models to have a higher Akaike information criterion (AIC) value compared to a model without the factor, the original stimulus factor was subsequently omitted as it did not contribute to better model fitting. Random slopes were by-participant and by-word pair for the effect of steps, and by-word pair on the effect of participant when deemed appropriate. For the *hana* pair, there was significant effect of the DLF group ($\chi^2(1) = 4.8$, p=0.029) and steps of the stimulus ($\chi^2(1) = 18.6$, p<0.001) on whether the listeners perceived the stimuli as "flower" or "nose". The *sita* pair however did not show any significant effect of the DLF group on the response ($\chi^2(1) = 1.3$, p=0.26) even though Fig. 2 shows the 'good' group had a distinct point where their perception changed from one to another at step 6 compared to the 'poor' group in and Fig. 3.

4. DISCUSSION

Our current study aimed to find out whether elderly listeners were able to perceive Japanese pitch accent accurately and how their perception was related to their frequency discrimination abilities. The results suggest that while some elderly listeners were able to identify the *hana* and the *sita* pair to a certain extent, others struggled with these words. As both groups of listeners had similar auditory thresholds and auditory filter bandwidths, their differences in word identification may be attributed to the difference in their pitch discrimination abilities.

Our second objective was to determine whether secondary cues, such as duration or intensity, for pitch accent in Japanese are present and accessible to elderly listeners, especially for those who may not be able to make use of the pitch cues. If secondary cues, whatever they are, were perceivable, we would see different results whether the stimuli were created from accented words or unaccented words. We can observe from Fig. 2 that this is not the case for the 'good' group, where they had almost identical responses for stimuli created from both original files. The 'poor' group, however, while there was no significant effect of the original stimulus, the results were visually more different. Having said that, the 'poor' group results also showed no evidence of being able to use any secondary cues, and therefore no acoustic analysis were carried out for the two original stimuli. F0, being the stronger acoustic cue, may have acted as a masker over any weaker secondary cues, and thus they were not available even to the listeners who could not use the pitch cues robustly.

Initially we expected the participants to perceive both word pairs similarly. Apart from one participant, who was trained as a broadcaster, the participants in general could not distinguish the *sita* pair as well as the *hana* pair. This may be because *hana*, which means "flower" when accented and "nose" when unaccented, are both concrete nouns that make sense somewhat in the carrier sentence and therefore easier for the listeners to perceive. On the other hand, "down" in the *sita* pair may not make sense contextually, and several participants reported afterwards to find the sentence awkward to comprehend. Previous studies have found that elderly listeners take advantage of context in adverse listening conditions [27, 14] and therefore have a bias towards the more semantically sensible choice when they could not decide. Phonologically, the *sita* pair may also be disadvantageous compared to *hana*, where the first mora in the *sita* stimulus is devoiced (confirmed via acoustic analysis), and therefore may be harder to hear. Previous literature has shown that elderly listeners struggle with devoiced sounds [17], possibly due to the tendency of higher frequency loss in elderly listeners. The mean hearing thresholds at 2k and 4k Hz for the current cohort were 18 and 19 dB HL respectively and thus can only be considered as mild hearing loss, suggesting that high frequency loss could not be used to explain the poor recognition for *sita*. While the stimuli were checked to sound natural and different enough, acoustically the f0 differences were smaller for the *sita* pair. This may have contributed to current results of elderly participants not being able to distinguish between the *sita* pair. Future work would involve further investigating whether this applies to young normal hearing listeners.

In regards to the effect of (or lack of) catathesis, we initially expected some sort of bias towards the unaccented choice of words, which was not observable in the results. For future study, stimuli with carrier phrase from the original accented words should also be included to test for listeners' sensitivity on the effect of catathesis.

There is also the possibility that elderly listeners were amusics to begin with and already had pitch processing deficits, where previous studies have found amusic listeners to have problems differentiating Mandarin tones [26, 19]. However, literature on amusics has found the mean relative DLF to be 5% [6], comparatively worse than our current cohort, suggesting that at least the participants of the current study had better pitch processing abilities than amusics. As we do not have data of the participants prior to their possible hearing decline, we cannot say their pitch processing deficits were a product of ageing. However, self reported experiences in music have shown that there were participants with music experience in both groups. For future work, we will also recruit young listeners to examine how they perceive pitch accent and whether they exhibit similar patterns to our elderly listeners.

Finally, all participants in the 'poor' group were male listeners. This may be due to our recordings being made from a female speaker and they may have found a male voice more familiar. At the same time, literature has also reported that speech intelligibility performances in male participants tend to be affected by ageing more than their female counterparts [8, 32] and the current results may be another piece of evidence for this observation.

5. CONCLUSION

In the current study, we examined how elderly Tokyo-born Japanese listeners perceive pitch accent using two pairs of accented and unaccented words: *hana* and *sita*. We found listeners who were similar in hearing thresholds and auditory filter bandwidths but differ in their frequency discrimination abilities to vary in how they perceive pitch accent in Japanese. The *hana* pair was more distinguishable than the *sita*, possibly due to the less natural context and less difference in pitch between the accented and unaccented pairs. We also used two sets of continua per word pair from the original accented and unaccented speech to investigate whether or not secondary cues may be present and accessible for listeners with poorer pitch processing abilities. As expected, the group with normal hearing DLF did not make use of possible secondary cues, and we did not find any evidence for the group with poorer pitch processing of having used possible secondary cues, even though visually the results suggested some differences. While we could not report conclusively the pitch deficits observed in the poorer group were due to ageing, we found evidence of poorer frequency discrimination abilities to affect how elderly listeners perceive day-to-day words. Future works involves using more word pairs and recruiting younger listeners to examine the effect of ageing.

6. REFERENCES

[1] Adank, P., Janse, E. 2010. Comprehension of a novel accent by young and older listeners. *Psychology and Aging* 25(3), 736–740.

[2] Amano, S., Kondo, T. 1999. *Nihongo no goitokusei*. Tokyo: Sanseido.

[3] Bates, D. M., Maechler, M., Bolker, B. 2015. lme4: Linear Mixed-Effects Models using 'Eigen' and S4.

[4] Beckman, M. 1986. *Stress and Non-Stress Accent*. Dordrecht: Foris.

[5] Boersma, P., Weenink, D. 2016. Praat: doing phonetics by computer. Available at http://www.fon.hum.uva.nl/praat/.

[6] Cousineau, M., Oxenham, A. J., Peretz, I. 2015. Congenital amusia: A cognitive disorder limited to resolved harmonics and with no peripheral basis. *Neuropsychologia* 66, 293–301.

[7] Cutler, A., Otake, T. 1999. Pitch accent in spoken-word recognition in Japanese. *J Acoust Soc Am* 105(3), 1877–1888.

[8] Dubno, J., Lee, F., Matthews, L., Mills, J. 1997. Age-related and gender-related changes in monaural speech recognition. *J Speech Lang Hear Res* 40, 444–52.

[9] Dubno, J. R., Dirks, D. D., Morgan, D. E. 1984. Effects of age and mild hearing loss on speech recognition in noise. *J Acoust Soc Am* 76(1), 87–96.

[10] Ellis, R. J., Molander, P., Rönnberg, J., Lyxell, B., Andersson, G., Lunner, T. 2016. Predicting Speech-in-Noise Recognition From Performance on the Trail Making Test. *Ear and Hearing* 37(1), 73–79.

[11] Fitzgibbons, P. J., Gordon-Salant, S. 1994. Age Effects on Measures of Auditory Duration Discrimination. *J Speech Hear Res* 37(3), 662.

[12] Füllgrabe, C., Moore, B. C., Stone, M. A. 2015. Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience* 7(JAN), 1–25.

[13] Gordon-Salant, S. 2006. Hearing loss and aging: New research findings and clinical implications. *The Journal of Rehabilitation Research and Development* 42(4), 9–24.

[14] Gordon-Salant, S., Fitzgibbons, P. 1997. Selected cognitive factors and speech recognition performance among young and elderly listeners. *J Speech Lang Hear Res* 40, 423–31.

[15] Gordon-Salant, S., Fitzgibbons, P. P. J. 1995. Comparing Recognition of Distorted Speech Using an Equivalent Signal-to-Noise Ratio Index. *J Speech Hear Res* 38(3), 706.

[16] Helfer, K., Wilber, L. 1990. Hearing loss, aging, and speech perception in reverberation and noise. *J Speech Hear Res* 33, 149–155.

[17] Iwagami, E., Arai, T., Yasu, K., Kobayashi, K. 2018. Misperception of Japanese words with devoiced vowels and/or geminate consonants by young and elderly listeners. *Acoustical Science and Technology* 39(2), 109–118.

[18] Kubozono, H. 1993. *The Organization of Japanese Prosody*. Tokyo: Kuroshio.

[19] Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., Stewart, L. 2012. The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers. *PLoS ONE* 7(2).

[20] Moore, B. C. 2014. *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss*. Singapore: World Scientific Publishing Company.

[21] Moore, B. C., Peters, R. W. 1992. Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *J Acoust Soc Am* 91(5), 2881–2893.

[22] Moore, B. C. J. 2016. Effects of sound-induced hearing loss and hearing AIDS on the perception of music. *AES: Journal of the Audio Engineering Society* 64(3), 112–123.

[23] Moore, B. C. J., Vickers, D. a., Mehta, A. 2012. The effects of age on temporal fine structure sensitivity in monaural and binaural conditions. *International Journal of Audiology* 51(10), 715–721.

[24] Nábělek, A. K., Robinson, P. K. 1982. Monaural and binaural speech perception in reverberation for listeners of various ages. *J Acoust Soc Am* 71(5), 1242–1248.

[25] Nakaichi, T., Watanuki, K., Sakamoto, S. 2003. A simplified measurement method of auditory filters for hearing-impaired listeners. *Acoustical Science and Technology* 24(6), 365–375.

[26] Nan, Y., Sun, Y., Peretz, I. 2010. Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain* 133(9), 2635–2642.

[27] Pichora-Fuller, M. K., Schneider, B. A., Daneman, M. 1995. How young and old adults listen to and remember speech in noise. *J Acoust Soc Am* 97(1), 593–608.

[28] Pierrehumbert, J. B., Beckman, M. E. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.

[29] Sugiyama, Y. 2017. Perception of Japanese Pitch Accent without F0. *Phonetica* 74(2), 107–123.

[30] Wang, Y., Yang, X., Zhang, H., Xu, L., Xu, C., Liu, C. 2017. Aging Effect on Categorical Perception of Mandarin Tones 2 and 3 and Thresholds of Pitch Contour Discrimination. *Am J Audiol* 26, 18–26.

[31] Weitzman, R. 1970. *Word Accent in Japanese*. Detroit: Whalen.

[32] Wiley, T. L., Cruickshanks, K. J., Nondahl, David, M., Tweed, T. S., Klein, R., Klein, B. E. K. 1998. Aging and word recognition in competing message. *J Am Acad Audiol* 9, 191–198.

[33] Wingfield, A., Poon, L., Lombardi, L., Lowe, D. 1985. Speed of processing in normal aging: effects of speech rate, linguistic structure, and processing time. *Journal of Gerontology* 40(5), 579–585.

[34] Yonan, C. A., Sommers, M. S. 2000. The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging* 15(1), 88–99.