Acoust. Sci. & Tech. **41**, 1 (2020)

ACOUSTICAL LETTER

Effect of frequency discrimination abilities on music and speech perception: A summary of four short studies

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 30 September 2018, Accepted for publication 28 May 2019*)

Keywords: Frequency discrimination, Speech perception, Interval recognition
PACS number: 43.71.+m, 43.75.Cd, 43.66.Sr, 43.66.Hg, 43.66.Fe [doi:10.1250/ast.41.408]

1. Introduction

The ability to discriminate pitch has been reported to decline with age in various studies (e.g. [1–3]). At the same time, we know that ageing affects speech perception, where age-related deficits are often studied together with hearing loss e.g. [4,5]. It would seem intuitive then that poor frequency discrimination would cause problems in speech and music perception, confirmed by Vongpaisal and Pichora-Fuller, who found age-related declines for F0 difference limens (FDLs) as well as identification of concurrently spoken vowels [2].

The current paper describes briefly four short pitch related studies on speech and music perception and highlights the differences between elderly listeners with good and poor frequency discrimination abilities.

The first study to be discussed is part of a bigger study on the perception of the long Japanese vowel word pair "obasan (aunt)" and "obaasan (grandmother)" [6]. The current summary will focus on the use of pitch to distinguish between the pair when duration becomes ambiguous. The second study looks at sentence intonation to distinguish between a question and a statement. The third study investigates the categorical perception of the Japanese vowels /o/ and /e/, where a continuum is created by manipulating the second formant. Finally in the fourth study the participants listen to four notes and determine whether they sound like the familiar melody of 'Twinkle Twinkle Little star.'

2. Methodology

2.1. Participants

Twelve elderly Japanese listeners participated in all four of the short studies, as well as tests to check their hearing thresholds, auditory filters, as well as pure tone frequency difference limens (FDLs) (geometric means for 8 out of 12 reversals) at three particular frequencies that correspond to the test stimuli for the long vowel and melody studies. The three frequencies were: 262 Hz (C4), 207 Hz (G#3) and 392 Hz (G4). The participants were then separated into two groups according to how well their frequency discrimination abilities were, where 7 of the participants performed lower than 20% relative FDL ($M = 6.64\%$, $SD = 5.25$), and 5 participants had over 20% relative FDL ($M = 35.93\%$, $SD = 15.54$). The relative FDL of the two groups were significantly different ($t(4.7) = -4.1, p = 0.01$).

The age of the participants range from 60 to 80 years old, with the mean age at 69.58. Their mean hearing threshold (average of 500 Hz, 1,000 Hz and 2,000 Hz) was 17.67 dB HL ($SD = 7.56$). The mean threshold at 250 Hz was 16.92 dB HL ($SD = 10.08$). There is no significant difference between the two groups in terms of hearing thresholds and age. Auditory filter bandwidths at 500 Hz, 1,000 Hz and 2,000 Hz were also measured but again there was no significant difference between the two groups.

For the long vowel study and the sentence experiments we also have two separate groups of young normal hearing listeners as control groups.

2.2. Test procedure

The speech stimuli were presented at 25 dB above their individual average hearing threshold at 500 Hz, 1,000 Hz and 2,000 Hz. The musical notes were presented at 25 dB above their hearing threshold at 250 Hz due to the lower register of the tune.

The participants 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-25E).

2.2.1. Long Japanese vowel word pair identification

Japanese differentiates between long and short vowels. While duration is obviously one of the auditory cues, previous research has found that pitch is being used as secondary cue when duration becomes ambiguous. In this study, the minimal pair "obaasan (grandmother)" and "obasan (aunt)" were chosen as stimuli as they differ in the length of /a/ and both words have similar familiarity for Japanese listeners. The participants listened to 36 stimuli both manipulated in duration and pitch within a carrier phrase to decide whether they regard the stimulus to mean 'aunt' or 'grandmother' and they repeated the test six times in random order. Each step of the pitch manipulation was approximately 85 cents. Only the pitch results for when duration (6 stimuli) is ambiguous will be reported in the current paper. This study has been published as part of a bigger study with the duration and pitch manipulation [6].

2.2.2. Sentence intonation

The sentence 'kegashita (injure-PST)' can mean 'Were you injured?' in question form, or 'I was injured' in statement form. The sentence intonation was manipulated in 10 steps to

*e-mail: justinehui@eagle.sophia.ac.jp
†e-mail: arai@sophia.ac.jp

create a continuum from question form to statement form and the participant judged whether the sentence is a statement or question ten times. The step between each stimulus was 85 cents.

2.2.3. Vowel quality

Japanese vowel formant values for /o/ and /e/ were taken from [7] isolated vowel data from 6 male speakers to create synthetic vowel sounds using the Klatt synthesier provided by [8]. A ten-step continuum was created by manipulating the second formant (F2) values from /o/ to /e/, using the average of the first and third formant values (F1 and F3) and a fundamental frequency (F0) of 150 Hz. The participants listened to the isolated vowels to identify whether they sound like /o/ or /e/ and repeated the test six times. The step between each F2 of the stimuli was 128 cents.

2.2.4. Familiar tune recognition

The participants listened to the first four notes from the musical tune 'Twinkle Twinkle Little Star,' where the first two notes and the following two notes are a perfect fifth apart. They were then asked to judge whether or not the tune sounds like 'Twinkle Twinkle' to them. The 3rd and 4th note are always the same tone, varying from 100 cents (1 semitone), 50, 40, 30, 20, 10 and 0 cent from the correct tone, in both directions, giving a total of 13 stimuli. Three registers were tested: C4 (440 Hz) - G4, F3 - C4 and G#3 - D#3 and the participants repeated the test 4 times, a total of 156 trials.

3. Results

The current paper will focus mainly on inspecting and comparing visually the results of the two groups (three groups including the young participants for the first two studies).

Figure 1 shows the results from the long vowel experiment when duration is considered to be ambiguous. The y-axis gives the average ratio of answers for participants choosing 'grandmother.' On the x-axis, we have the pitch levels, where P0 resembles the pitch contour for 'obasan (aunt)' and P5 as 'obaasan (grandmother).' We observe that the group with good frequency just-noticeable-difference (JND) is able to use pitch as a cue to distinguish between 'aunt' and 'grandmother' when duration becomes unreliable, displaying similar results to the young group. On the other hand, the group with poor FDLs could not use pitch cues reliably to differentiate between the word pair.

For the sentence experiment, the y-axis shows the average ratio of responses for 'statement,' where in the x-axis, 0 step resembles a question, and 10 step resembles a statement. While we see less of a steep s-curve even in the good group compared to the young group, it is clear that the bad group hovers around chance (0.5), despite some recognition towards the statement end in Fig. 2.

In the vowel identification test, we have two s-curves from identifying the stimulus as /o/ at stimulus 0, to /e/ at stimulus 10, as shown in Fig. 3. Both groups were able to identify the vowels at the two extremes, but we still see a difference between the groups, with the poor group perceiving 'e' later on the continuum.

Lastly, Fig. 4 shows the results from the familiar tune test. The ideal correct response would be average ratio of 'Yes' as answer at 1 when step is at 0 cent, and 0 at all other steps. We see here that the good group can perform the task relatively well, with a curve centring around 0 (slightly towards the sharper end) and curving down to 0 on both ends. The poor group on the other hand hovers around chance (0.5), suggesting that they could not make out the differences between the stimuli.

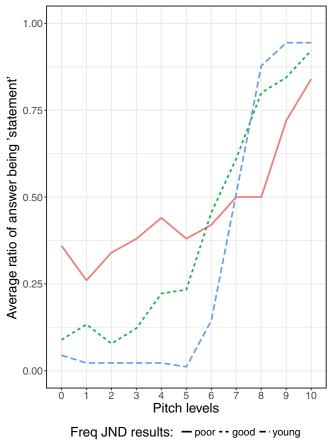

Fig. 2 Sentence type.

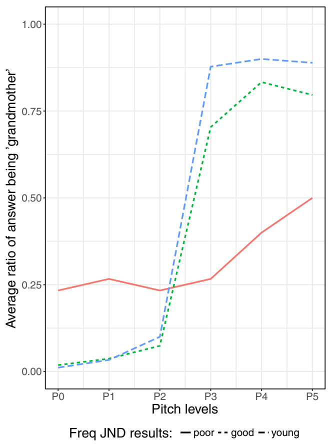

Fig. 1 Long vowel.

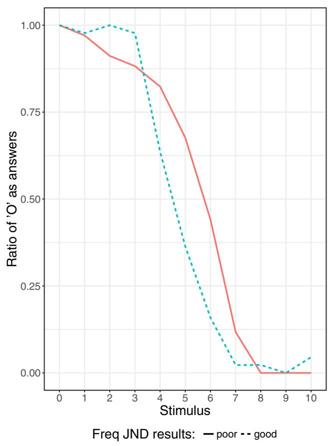

Fig. 3 Vowel quality.

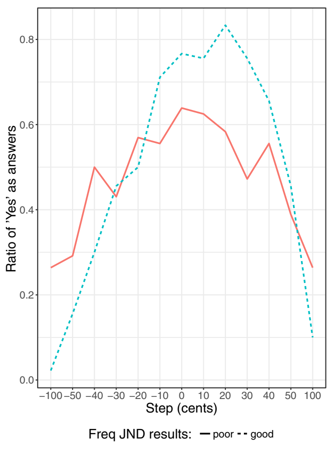

Fig. 4 Familiar melody.

4. Discussion

It seems logical that with poorer frequency discrimination, comes poorer speech and music perception. We can see in the current paper from the four short experiments that even if the participants had similar thresholds, auditory filter bandwidth and age, there can be a big difference in how they perceive speech and music depending on their ability to discriminate frequency.

In the current study, our 'poor' group with near-normal hearing scored an average of 35% relative FDLs, more than two times that of previous studies, where elderly listeners with hearing loss were reported to have around 10% relative FDLs [1]. The participants in the current study were made sure to understand the task and were later interviewed, reporting the tones to sound the same. The unusually high relative FDLs may be of interest in itself, and may suggest why some elderly listeners have hearing problems despite near-normal thresholds. This also supports the findings from [5] that auditory thresholds do not correlate with FDLs, where it was reported that normal hearing elderly listeners had the worst FDLs, even more so than hearing impaired elderly listeners.

There is also the possibility of these participants being amusics already, and have difficulty with pitch recognition regardless of their age. However, the average relative FDLs for young amusics were found to be around 5% [9], which is still much less than the poor group in the current study. Previous studies have shown that while amusic participants have problems with music perception, it does not usually affect speech perception [10]. In the current study we found that not only does a poor FDL affect music interval perception, there also seems to be a change in vowel quality, to which we do not have a satisfactory explanation.

5. Conclusion

In the current paper, we compared two groups of near-normal hearing elderly listeners with no significant difference in hearing thresholds and auditory filter bandwidths, but differ in frequency difference limens. The results from four small studies showed that poor FDLs affect speech on the word, sentence and vowel level, as well as musical interval perception.

References

[1] 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).

[2] T. Vongpaisal and M. K. Pichora-Fuller, "Effect of age on F0 difference limen and concurrent vowel identification," *J. Speech Lang. Hear. Res.*, **50**, 1139–1156 (2007).

[3] 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).

[4] S. Gordon-Salant, "Hearing loss and aging: New research findings and clinical implications," *J. Rehabil. Res. Dev.*, **42**(4), pp. 9–24 (2006).

[5] M. Coughlin, D. Kewley-Port and L. E. Humes, "The relation between identification and discrimination of vowels in young and elderly listeners," *J. Acoust. Soc. Am.*, **104**, 3597–3607 (1998).

[6] C. T. J. Hui and T. Arai, "Elderly listeners' identification of Japanese long vowel pair 'obasan' and 'obaasan' using pitch and duration," *Acoust. Sci. & Tech.*, **40**, 105–115 (2019). URL: https://www.jstage.jst.go.jp/article/ast/40/2/40_E1825/_article

[7] H. Fujisaki and M. Sugito, "Onsei no Butsuri-teki seishitsu," in *Iwanami Koza, Nihongo 5 On'in* (Iwanami Shoten, Tokyo, 1977), pp. 63–106.

[8] H. T. Bunnell, Simplified Vowel Synthesis Interface (2015). URL: http://www.asel.udel.edu/speech/tutorials/synthesis/vowels.html

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

[10] J. Ayotte, I. Peretz and K. Hyde, "Congenital amusia: A group study of adults afflicted with a music-specific disorder," *Brain J. Neurol.*, **125**(Pt 2), 238–251 (2002).