RATE DISCRIMINATION, SENTENCE AND PROSODY RECOGNITION IN YOUNG AND ELDERLY NORMAL HEARING ADULTS USING VOCODERS

UCONN

Lendra M. Friesen[1] , Robert P. Morse[1], Monita Chatterjee[2]

[1]University of Connecticut, Storrs, CT, USA, [2]Boys Town National Research Hospital, Omaha, NE, USA

BOYS TOWN National Research Hospital

lendra.friesen@uconn.edu

Conference on Implantable Auditory Prostheses, July 14-19, 2019, Lake Tahoe, CA

Introduction

Temporal-envelope based voice-pitch coding is important for listeners with hearing impairment, especially listeners with cochlear implants (CIs), as spectral resolution is not sufficient to provide a spectrally based voice-pitch cue. More specifically, the encoding of fundamental frequency (F0) information is critical to communicating in such environments. Voiced speech can be approximated by a harmonic complex, with the perceived pitch of listeners' voices roughly corresponding to the F0 of the harmonic complex. Voiced pitch information has also been used in other types of processing such as speech intonation recognition (Lehiste, 1970), lexical tone recognition (Chao, 1968), and talker-gender identification (Titze,1989), but also to separate competing sound sources (Brokx and Nootebohm, 1982; Brungart, 2001) and to determine speaker authenticity and voice emotion (Drolet et al., 2014).

Differences in F0 discrimination between young and elderly normal hearing individuals has been shown, with elderly individuals performing more poorly (Vongpaisel & Pichora-Fuller, 2007). This difference has also been shown to occur in young and elderly individuals listening through CI vocoders with different numbers of spectral channels (Schvartz-Leyzac & Chatterjee, 2015). However, it is still somewhat unknown as to how this encoding relates to more real-life listening situations such as speech understanding in quiet and noise and prosody recognition.

Aims

The objectives of this study are to compare results between younger and elderly individuals with normal hearing (at least up to 2000 Hz) and measure noise-vocoded listening by:

1) Measuring F0 discrimination using 8 and 24 channel vocoders, with a 2000 and 4000 Hz cut-off
2) Measuring 8-channel sentence recognition in quiet and in +5 SNR babble
3) Measuring prosody recognition with the 8-channel vocoder

AIM I: Is there a difference in F0 discrimination when listening to vocoders between young and elderly listeners?

Participants:

- 5 young (22-26 years) NH listeners (≤ 20 dB HL thresholds from 250 to 8000 Hz; normal tympanograms)
- 5 elderly (65-78 years) NH listeners (≤ 20 dB HL thresholds from 250 to 2000 Hz; normal tympanograms)
- All participants scored within the normal range on the Montreal Cognitive Assessment (MoCA)

Stimuli:

- Signals were created online (44100Hz sampling rate) and delivered through a custom graphical user interface developed in MATLAB
- Stimuli were harmonic complexes,400ms in duration
- The noise band vocoding procedure was similar to Shannon et al. (1995)
- The F0 and partials were summed to create a harmonic complex series similar to Schvartz-Leyzac & Chatterjee (2015)
- However 4 conditions were created:
    - Equal-amplitude harmonics between 100-2000 Hz or 100-4000 Hz and with each harmonic series bandpass-filtered into 8 or 24 channels

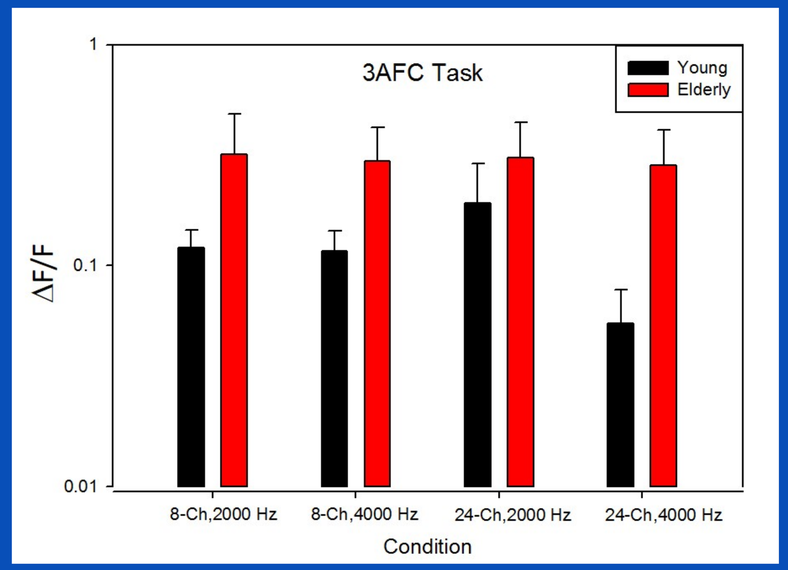
Procedure:

- Two-down, one-up, 3-AFC procedure was used to measure F0 discrimination (threshold 70.7%; Levitt, 1971)
- Two of the intervals contained a reference stimulus with F0 value equal to 100 Hz, while a third interval contained the experimental value (always greater than the reference F0)
- The experimental stimulus was presented at random in one of the three intervals
- The F0 value of the experimental stimulus was adapted for a maximum of ten reversals or 55 total trials (whichever occurred first)
- A minimum of eight reversals was required to calculate the average F0 value; if eight reversals could not be reached in 55 trials, the run was aborted
- Initial and final adaptive step sizes varied depending on the condition and listener's sensitivity, but were generally 4 and 2 Hz
- The mean was calculated from the last four reversals of a run, and that value was taken as the F0 difference limen
- Stimuli were output through an external soundcard [Edirol 25-UAEX (Roland Corporation US, Los Angeles,CA)] and mixer [RaneSM26B (Rane Corporation, Mukilteo,WA)], before being delivered through calibrated insert earphones
- All stimuli had equal rms values and were delivered binaurally at approximately 65 dBA

Procedure Cont'd:

- Participants were tested in a double-walled, sound-proof booth using a computer interface
- The computer interface displayed boxes (labeled "1," "2," and "3") that lit up simultaneously with the sequential presentation of the corresponding reference or experimental stimuli
- Listeners were asked to use a mouse to click on the box that contained the "different" sound (experimental F0)
- The inter-stimulus-interval was 400 ms, and participants were given unlimited time to respond
- After the participant made a selection, the next sequence of sounds was played 600 ms later.
- Participants received practice before being tested and received feedback on the display after each response
- After completing the practice runs, listeners completed at least two test trials for each condition in random order
- These two runs were used to calculate the final mean performance. When the difference between the thresholds of the two means was greater than 15%, another run was performed, and then the average of all three runs was calculated as the final mean.

AIM 1 RESULTS



There was a significant main effect for age $F(1,4) = 10.8$, $p = .03$; however, neither channel number or cut-off frequency revealed significant results.

AIM 2: Is there a difference between young and elderly listeners when measuring sentence recognition with 8-channel vocoded F0 harmonic complexes in quiet and in noise?

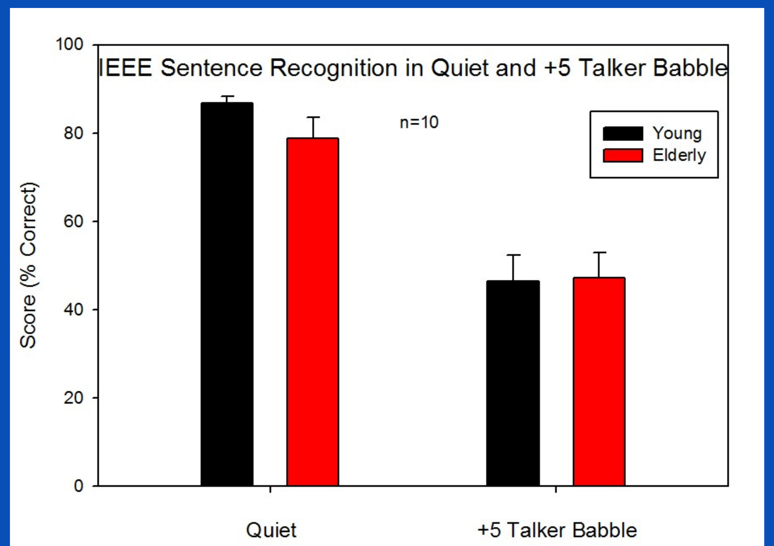
Stimuli:

- IEEE sentences processed in similar manner to AIM 1, using 8 channels and 100 Hz F0 start and 2000 Hz cut-off
- The speech was presented at 65 dBA using a custom written JAVA program

Procedure:

- Stimuli were output through an external soundcard [Edirol 25-UAEX (Roland Corporation US, Los Angeles,CA)] and mixer [RaneSM26B (Rane Corporation, Mukilteo,WA)], before being delivered binaurally through calibrated insert earphones
- One list of 10 processed sentences was presented as a training test
- Two lists of 10 sentences in quiet and 8-talker babble at a +5 SNR were presented
- Lists were scored in terms of % correct

AIM 2 Results



There was a significant main effect for noise $F(1,4) = 258$, $p < .001$; however, there was not a significant effect for age.

AIM 3: Is there a difference between young and elderly listener prosody recognition using 8-channel F0 vocoders?

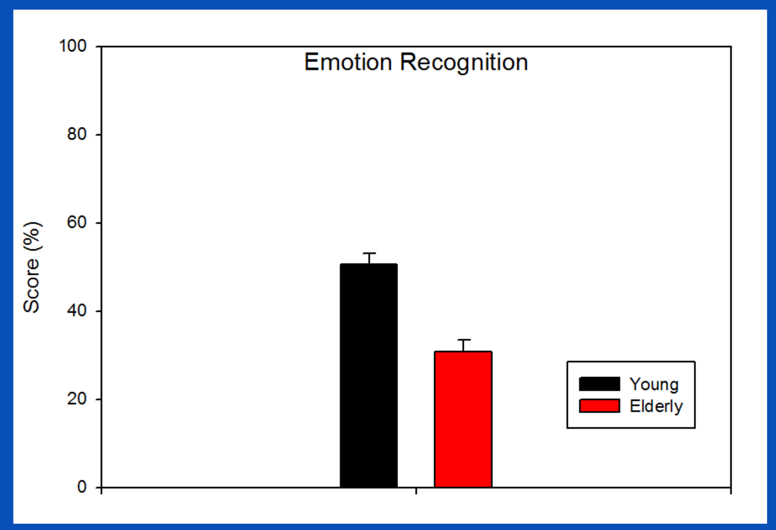
Stimuli:

- Were from the House Ear Institute emotional speech database (HEI-ESD) (Luo, 2007)
- 1 male talker produced 50 semantically neutral, everyday English sentences according to 5 target emotions (angry, happy, sad, anxious, and neutral)
- Presented with i-CAST computer software (Tigerspeech Technology, 2008)

Procedure:

- Stimuli were output through an external soundcard [Edirol 25-UAEX (Roland Corporation US, Los Angeles,CA)] and mixer [RaneSM26B (Rane Corporation, Mukilteo,WA)], before being delivered binaurally through calibrated insert earphone
- 100 sentences in the 5 target emotions were presented at 65 dBA
- Lists were scored in terms of % correct

Aim 3 Results



An independent t-test revealed a significant effect $t(8) = 5.8$, $p < .001$.

Summary

- Significant main effect of age in the 3AFC task; however, no significant effects for channel number and cut-off filter
- Significant main effect of listening environment in the sentence recognition task, but not for age
- Significant main effect for age in the emotion recognition task

Concluding Remarks

- The significant difference in F0 discrimination between the young and elderly groups could be at least one of the reasons for the difference in emotion recognition between these same groups.
- More participant data needs to be collected.
- Electrophysiology results will be added to help determine where the neural generators are located in the young and elderly listeners.

References

•Brokx, J. P. L., and Nootebohm, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics 10, 23–36.
•Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. 109,1101–1109.
•Chao, Y. R. (1968). "A Grammar of Spoken Chinese", (University of California Press, Berkeley, CA), 847 pp.
•Drolet, M., Schubotz, R. I., and Fischer, J. (2014). "Recognizing the authenticity of emotional expressions: F0 contour matters when you need to know," Front. Human Neurosci. 8(144), 1–11.
•Lehiste, I. (1970). "Suprasegmental features of speech," in Contemporary Issues in Experimental Phonetics, edited by N. J. Lass (Academic Press, New York), pp. 225–239.
•Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. 49, 467–477.
•Luo, X., Fu, Q. J., & Galvin, J. J. (2007). "Cochlear Implants Special Issue Article: Vocal Emotion Recognition by Normal-Hearing Listeners and Cochlear Implant Users", *Trends in Amplification*. https://doi.org/10.1177/1084713807305301
•Vongpaisal, T., & Pichora-Fuller, M. K. (2007). "Effect of Age on F 0 Difference Limen and Concurrent Vowel Identification", *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2007/079)
•Shannon, R. V., Zeng, F-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," Science 279, 303–304.
•Schvartz-Leyzac, K. C., & Chatterjee, M. (2015). "Fundamental-frequency discrimination using noise-band-vocoded harmonic complexes in older listeners with normal hearing", .J. Acoust. Soc. Am. https://doi.org/10.1121/1.4929938

Acknowledgements

The authors wish to thank the subjects who participated in these experiments for their time and effort. Without them we could not have completed this project.