

Research Article

Do Older Listeners With Hearing Loss Benefit From Dynamic Pitch for Speech Recognition in Noise?

Jing Shen^a and Pamela E. Souza^a

Purpose: Dynamic pitch, the variation in the fundamental frequency of speech, aids older listeners' speech perception in noise. It is unclear, however, whether some older listeners with hearing loss benefit from strengthened dynamic pitch cues for recognizing speech in certain noise scenarios and how this relative benefit may be associated with individual factors. We first examined older individuals' relative benefit between natural and strong dynamic pitches for better speech recognition in noise. Further, we reported the individual factors of the 2 groups of listeners who benefit differently from natural and strong dynamic pitches.

Method: Speech reception thresholds of 13 older listeners with mild-moderate hearing loss were measured using target speech with 3 levels of dynamic pitch strength.

Individuals' ability to benefit from dynamic pitch was defined as the speech reception threshold difference between speeches with and without dynamic pitch cues.

Results: The relative benefit of natural versus strong dynamic pitch varied across individuals. However, this relative benefit remained consistent for the same individuals across those background noises with temporal modulation. Those listeners who benefited more from strong dynamic pitch reported better subjective speech perception abilities.

Conclusion: Strong dynamic pitch may be more beneficial than natural dynamic pitch for some older listeners to recognize speech better in noise, particularly when the noise has temporal modulation.

This special issue contains papers from the 2016 Hearing Across the Lifespan (HEAL) conference held in Cernobbio, Italy.

Dynamic pitch, the natural variation in fundamental frequency, is one of the major prosodic cues in speech. In clear speech, dynamic pitch facilitates processing of speech, such as word segmentation (e.g., Brown, Salverda, Dille, & Tanenhaus, 2011; Cutler, 1976), complex syntax (e.g., Weber, Grice, & Crocker, 2006), and affective information (Fairbanks, 1940; Frick, 1985). We know that natural dynamic pitch helps younger listeners recognize speech better in noisy environments (e.g., Binns & Culling, 2007; Laures & Bunton, 2003). Our recent data further demonstrated this effect of dynamic pitch on speech perception in noise with a group of older listeners with a wide range of hearing status (Shen & Souza, 2017).

As described by our earlier article (Shen & Souza, 2017), older listeners as a group perceive speech better

with natural dynamic pitch than with strong dynamic pitch. The individual variability, however, was substantial in this data set. Recognizing speech in noise is difficult for older listeners with hearing loss. In an effort to help these listeners benefit from dynamic pitch, we first tested the hypothesis that strong dynamic pitch may provide extra help for some individuals. This hypothesis is grounded on three accounts. First, for those older listeners who have substantial difficulty hearing speech in noise, any extra perceptual cues that may be redundant for their peers could provide additional benefit (Bernstein & Grant, 2009). Second, we know that older listeners rely heavily on prosodic cues—including dynamic pitch—for speech comprehension in a quiet environment (Wingfield, Wayland, & Stine, 1992; see also Wingfield & Tun, 2001, for a review). Stronger dynamic pitch cues can provide more salient prosody and, therefore, be more beneficial in processing speech information. Last and most important, substantial variability in the benefit from strengthened dynamic pitch has been demonstrated by previous research (Grant, 1987; Miller, Schlauch, & Watson, 2010). Grant (1987) measured identification of pitch contour continua and observed large individual variability among the five listeners with hearing loss. Three of them were able to identify pitch contours that were slightly stronger (one to two

^aDepartment of Communication Sciences and Disorders, Northwestern University, Evanston, IL

Correspondence to Jing Shen: jing.shen@northwestern.edu

Editor-in-Chief: Sumitrajit Dhar

Received December 30, 2016

Revision received June 5, 2017

Accepted June 19, 2017

https://doi.org/10.1044/2017_AJA-16-0137

Disclosure: The authors have declared that no competing interests existed at the time of publication.

times) than normal contours. For the other two listeners, pitch contours were only useful when they were three to six times stronger than the natural ones. This variability was also associated with individuals' ability to perceive pitch information in speech (i.e., intonation and stress contrast). Focusing on the individual variability, the primary goal of our article is to examine whether older listeners with hearing loss vary in the dynamic pitch strength for achieving best speech recognition performance in noise. In other words, is stronger dynamic pitch more beneficial for some listeners, whereas natural dynamic pitch works better for others?

Second, while listeners benefit from dynamic pitch in a variety of noise conditions (Binns & Culling, 2007; Laures & Bunton, 2003), there is currently no data of individuals' dynamic pitch benefit compared across different noises. Specifically, we do not know if an individual's relative benefit pattern holds across noises with different levels of temporal modulation. Following the findings that individuals' ability to perceive speech in temporally modulated noise is relatively consistent across noise conditions (e.g., Bacon, Opie, & Montoya, 1998), it is expected that if a listener benefits more from strong dynamic pitch in highly modulated noise, he or she will also show the same benefit pattern in a different noise scenario that has temporal modulation. If this result is observed, it would also suggest a glimpsing mechanism for dynamic pitch benefit, in which older individuals benefit more easily from pitch cues when they have access to these cues through momentary dips in noise.

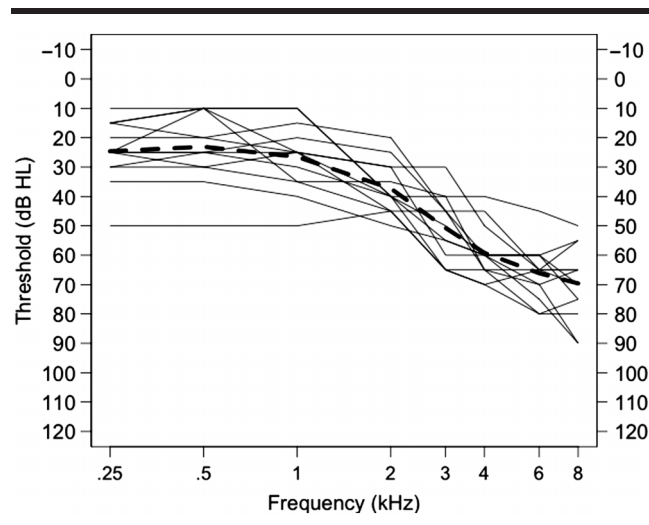
Further, we are interested in the characteristics of those listeners who benefit more from strong dynamic pitch than the natural ones, as this information may be useful for development of future speech-in-noise treatment. For instance, if we are going to implement a signal-processing strategy to strengthen dynamic pitch, who is likely to benefit from this enhancement? In Grant (1987), the five listeners' performance could not be explained by audiometric testing results; with a group that has less hearing loss but more advanced age than Grant's sample, we explore whether relative benefit from dynamic pitch is related to individual factors, such as age, audiometric profile, and subjective speech perception ability.

Method

Participants

Thirteen older adults (four women and nine men) aged 60–83 years (mean age = 73.5 years) participated in this study. Participants had mild–moderate sensorineural hearing loss as defined by pure-tone threshold > 25 dB HL (American National Standards Institute, 2009) between 0.25 kHz and 2.00 kHz and/or > 35 dB HL at 3.00 kHz. Figure 1 shows audiograms for the 13 listeners. While all had mild-to-moderate sensorineural hearing loss, they varied slightly in hearing loss degree and slope. All participants were native English speakers with no experience in tonal languages, had no or minimal musical experience, and passed a screening for mild cognitive impairment.

Figure 1. Audiogram of the listeners ($N = 13$, individual thresholds: thin lines, group average: thick dashed line).



Stimuli and Procedure

The target speech stimuli were drawn from Harvard IEEE sentences (Rothauser et al., 1969) that were produced by a female talker. The fundamental frequency contours of the stimuli were manipulated, and the sentences were resynthesized using PRAAT (Boersma & Weenink, 2013). The pitch contour of the speech was manipulated using the following formula:

$$\begin{aligned} \text{Instant } f_0 &= \text{Sentence average } f_0 + (\text{Original} \\ &\quad \text{instant } f_0 - \text{Sentence average } f_0) \\ &\quad \times \text{Pitch factor.} \end{aligned} \quad (1)$$

The pitch factor was set to 0 in the monotone condition, 1 in the original pitch condition, and 1.75 in the strong pitch condition. The International Collegium for Rehabilitative Audiology (ICRA) noises (Dreschler, Verschuure, Ludvigsen, & Westermann, 2001) were used as background noises, with three levels of temporal modulation: unmodulated one-talker speech-shaped noise, modulated one-talker speech-shaped noise, and modulated two-talker speech-shaped noise.

Prior to experimental testing, participants completed an audiometric battery consisting of case history, otoscopy, pure-tone threshold testing, and word recognition in quiet with Northwestern University Auditory Test No. 6 25-word lists (Tillman & Carhart, 1966). Subjective speech perception ability was measured using the Speech, Spatial and Qualities of Hearing Scale questionnaire (Gatehouse & Noble, 2004), which emphasizes specific difficult listening situations, including noisy background and multiple talker scenarios. It contains three domains: speech recognition, spatial aspect of hearing, and qualities of hearing experience.

Speech reception thresholds (SRTs) were obtained for all nine conditions (3 Dynamic Pitch Conditions \times 3 Background Noise Conditions) with an adaptive procedure

(Plomp & Mimpfen, 1979). In this paradigm, the signal-to-noise ratio was changed trial by trial on the basis of the listener's response to track the performance level of 50% correct. The output stimuli were amplified using the National Acoustics Laboratories–Revised linear prescriptive formula on the basis of individual thresholds (Byrne, Dillon, Ching, Katsch, & Keidser, 2001). Stimuli were presented monaurally in the better ear at 68 dB SPL using an M-Audio FastTrackPro external soundcard (M-Audio, Cumberland, RI) and an ER-2 insert earphone (Etymotic Research, Elk Grove, IL) in the test ear. Listeners were seated in a double-wall sound-proofed booth and were instructed to repeat back the sentences aloud for the experimenter, who scored the responses from outside of the booth. Individuals' benefits from dynamic pitch cues were defined as the difference between the SRTs in the monotone condition as compared with the original or strengthened pitch conditions.

Results and Discussion

Among the 13 older listeners, the average amount of benefit each individual had from dynamic pitch cues spread across a range of 1.2 dB to 5.2 dB signal-to-noise ratio, which translates to a 40% range in percentage of correct responses in a speech recognition in noise task (e.g., Festen & Plomp, 1990). On an individual level, some older listeners recognized speech better with the natural dynamic pitch as compared with the strong one; others performed better with the strong dynamic pitch. Figure 2 presents individual data of benefit from natural and strong dynamic pitches in three noise scenarios, respectively. It shows the magnitude of the dynamic pitch benefit, as well as the scattered pattern of the relative benefit across individuals.

As to the individuals' relative benefit across noise conditions, there was a significant correlation between the relative benefit scores (as indicated by the difference between

SRT with original and SRT with strong dynamic pitch contours) in strongly and mildly modulated noises, $r = .53$, $t(11) = 2.12$, $p = .05$, which suggests that relative benefit is consistent for the same individual across the two temporally modulated noise conditions. That is, individuals who benefited more from strong dynamic pitch in highly modulated noise performed in a similar way in mildly modulated noise. This pattern, however, did not transfer to the unmodulated noise condition, as the relative benefit scores were not correlated between this noise scenario and the modulated noises ($p > .1$). Overall, these results suggest that the individuals' relative benefit from dynamic pitch for speech recognition is consistent across noise scenarios with similar temporal characteristics. This finding is also consistent with the speech-in-noise literature (e.g., George, Festen, & Houtgast, 2006) in highlighting the importance of temporal modulation in noise to speech recognition, as it provides momentary access to speech signal—and to pitch cues in this case.

When the 13 listeners were grouped based on the relative benefit from dynamic pitch (collapsed over the two modulated noises), one third of them (five among 13 listeners) had more benefit from strong dynamic pitch than the natural ones, and the rest (eight) had the opposite pattern. The individual factors of the two groups are reported in Table 1 (standard deviations in parentheses). They were comparable in age and high-frequency pure-tone average. The strong pitch group had more hearing loss as indicated by higher pure-tone averages, but this difference was not statistically significant, $t(7) = -1.68$, $p = .1$. The subjective speech perception scores were analyzed using Wilcoxon rank-sum tests. The two groups were comparable in two Speech and Spatial Qualities subscales (Spatial: $W = 27$, $p = .34$, Quality: $W = 21$, $p = .94$), but the strong pitch group had higher scores (i.e., better performance) on the Speech subscale, which is marginally significant ($W = 8$, $p = .09$). This means that those listeners who were able to benefit

Figure 2. Benefit scores of individual participants (x axes: original dynamic pitch benefit; y axes: strong dynamic pitch benefit). Each dot represents one individual, and each panel is a noise condition. Individuals who are located above the diagonal lines benefit more from strong dynamic pitch than from original dynamic pitch; those below the diagonal lines benefit more from original dynamic pitch. ICRA = International Collegium for Rehabilitative Audiology; SNR = signal-to-noise ratio.

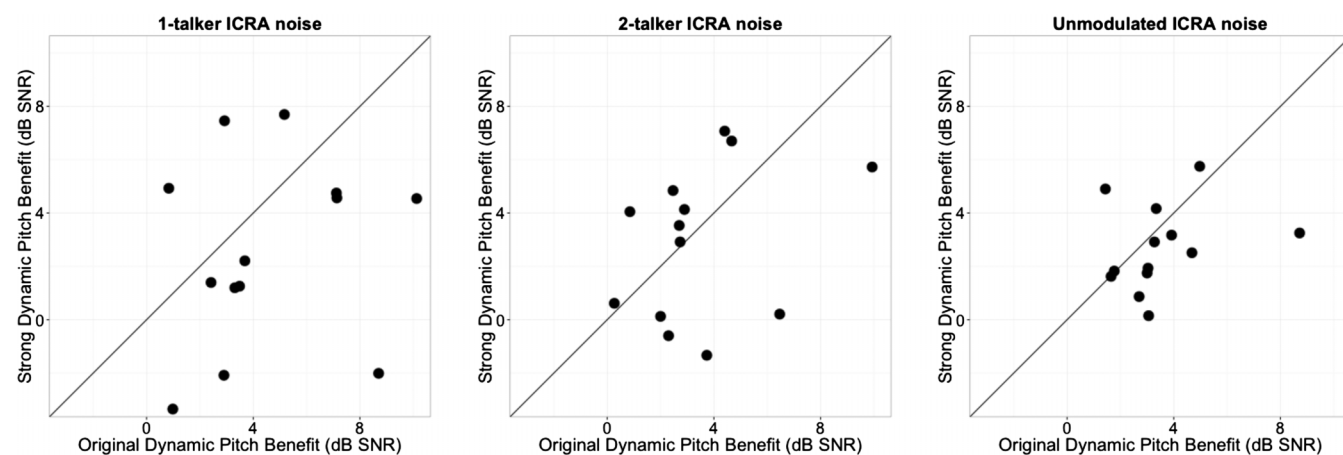


Table 1. Comparison of the individual factors between the two groups of listeners.

Group	Mean age (years)	PTA (0.5 kHz, 1 kHz, 2 kHz, dB HL)	High-frequency PTA (3 kHz, 4 kHz, 6 kHz, dB HL)	SSQ Speech Scale mean score	SSQ Spatial Scale mean score	SSQ Quality Scale mean score
Natural pitch group (<i>n</i> = 8)	73.6 (7.19)	25.8 (7.07)	58.9 (7.31)	6.31 (1.78)	7.98 (0.82)	7.78 (1.25)
Stronger pitch group (<i>n</i> = 5)	73.5 (4.47)	34.0 (8.13)	58.3 (8.36)	7.40 (2.15)	6.91 (1.44)	7.87 (1.04)

Note. Natural pitch group benefit more from natural than stronger dynamic pitch; stronger pitch group benefit more from stronger than natural dynamic pitch. PTA = pure-tone average; SSQ = Speech and Spatial Qualities questionnaire.

from strong dynamic pitch cues (as compared with natural ones) tended to do better in difficult scenarios, such as with multiple talkers and/or background noise, and this advantage could not be attributed to hearing sensitivity. This finding aligns with the previous research on stream segregation (e.g., Mackersie, Prida, & Stiles, 2001) in suggesting the role of individuals' perceptual abilities, which may not be tightly coupled with hearing sensitivity, in their speech perception in complex listening scenarios. Although these results were limited by the small sample size, this data set nevertheless revealed an interesting direction that merits further investigation.

The present study served as a first step in this line of research with the ultimate goal of devising individualized clinical remedies that make dynamic pitch helpful for older listeners with hearing loss to perceive speech better in noise. The finding can be informative for the development of advanced assistive hearing technology for enhancing dynamic pitch cues on the basis of individuals' perceptual pattern. Further research is warranted to investigate this individualized dynamic pitch benefit using a fine-grained scale of pitch strength, as well as including other patient factors that may contribute to the relative benefit.

Acknowledgments

This work was supported by the National Institutes of Health (Grants F32DC014629, awarded to Jing Shen, and R01DC12289, awarded to Pamela Souza). The authors thank Richard Wright for helpful suggestions on the study design; Arleen Li, Laura Mathews, and Paul Reinhart for assistance with data collection; and Tim Schoof for help with the experiment program. The data were presented at the Hearing Across the Lifespan Conference, Cernobbio, Italy.

References

- American National Standards Institute. (2009). *American National Standards Institute Specification of Hearing Aid Characteristics*. New York: ANSI.
- Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal of Speech, Language, and Hearing Research*, 41(3), 549–563.
- Bernstein, J. G., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 125(5), 3358–3372.
- Binns, C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *The Journal of the Acoustical Society of America*, 122(3), 1765–1776.
- Boersma, P., & Weenink, D. (2013). PRAAT: Doing phonetics by computer (Version 5.3.82) [Computer software]. Retrieved from <http://www.fon.hum.uva.nl/praat/>
- Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review*, 18(6), 1189–1196.
- Byrne, D., Dillon, H., Ching, T., Katsch, R., & Keidser, G. (2001). NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures. *Journal of the American Academy of Audiology*, 12(1), 37–51.
- Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, 20(1), 55–60.
- Dreschler, W. A., Verschuure, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *International Journal of Audiology*, 40(3), 148–157.
- Fairbanks, G. (1940). Recent experimental investigations of vocal pitch in speech. *The Journal of the Acoustical Society of America*, 11, 457–466.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88(4), 1725–1736.
- Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, 97(3), 412.
- Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, 43(2), 85–99.
- George, E. L., Festen, J. M., & Houtgast, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 120(4), 2295–2311.
- Grant, K. W. (1987). Identification of intonation contours by normally hearing and profoundly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 82, 1172–1178.
- Laures, J. S., & Bunton, K. (2003). Perceptual effects of a flattened fundamental frequency at the sentence level under different listening conditions. *Journal of Communication Disorders*, 36(6), 449–464.
- Mackersie, C. L., Prida, T. L., & Stiles, D. (2001). The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, 44(1), 19–28.

- Miller, S. E., Schlauch, R. S., & Watson, P. J. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. *The Journal of the Acoustical Society of America*, 128(1), 435–443.
- Plomp, R., & Mimpfen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America*, 66(5), 1333–1342.
- Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., & Weinstock, M. (1969). I.E.E.E. recommended practice for speech quality measurements. *IEEE Trans. Audio Electroacoust.*, 17(3), 225–246.
- Shen, J., & Souza, P. (2017). The effect of dynamic pitch on speech recognition in temporally modulated noise. *Journal of Speech, Language, and Hearing Research*. Advance online publication. https://doi.org/10.1044/2017_JSLHR-H-16-0389
- Tillman, T. W., & Carhart, R. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6*. Evanston, IL: Auditory Research Lab, Northwestern University.
- Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, 99(2), B63–B72.
- Wingfield, A., Wayland, S. C., & Stine, E. A. (1992). Adult age differences in the use of prosody for syntactic parsing and recall of spoken sentences. *Journal of Gerontology*, 47(5), P350–P356.
- Wingfield, A., & Tun, P. A. (2001). Spoken language comprehension in older adults: Interactions between sensory and cognitive change in normal aging. *Seminars in Hearing*, 22, 287–302.