JSLHR

Research Article

The Effect of Dynamic Pitch on Speech Recognition in Temporally Modulated Noise

Jing Shen[a] and Pamela E. Souza[a]

Purpose: This study investigated the effect of dynamic pitch in target speech on older and younger listeners' speech recognition in temporally modulated noise. First, we examined whether the benefit from dynamic-pitch cues depends on the temporal modulation of noise. Second, we tested whether older listeners can benefit from dynamic-pitch cues for speech recognition in noise. Last, we explored the individual factors that predict the amount of dynamic-pitch benefit for speech recognition in noise.

Method: Younger listeners with normal hearing and older listeners with varying levels of hearing sensitivity participated in the study, in which speech reception thresholds were measured with sentences in nonspeech noise.

Results: The younger listeners benefited more from dynamic pitch for speech recognition in temporally modulated noise than unmodulated noise. Older listeners were able to benefit from the dynamic-pitch cues but received less benefit from noise modulation than the younger listeners. For those older listeners with hearing loss, the amount of hearing loss strongly predicted the dynamic-pitch benefit for speech recognition in noise.

Conclusions: Dynamic-pitch cues aid speech recognition in noise, particularly when noise has temporal modulation. Hearing loss negatively affects the dynamic-pitch benefit to older listeners with significant hearing loss.

Pitch, as defined by the percept of fundamental frequency (f0) in speech, is one of the most powerful cues for speech recognition. In a quiet environment, pitch conveys important linguistic information for the perception of speech information, including phonemes and words (e.g., Faulkner & Rosen, 1999; Holt, Lotto, & Kluender, 2001; Spitzer, Liss, & Mattys, 2007; Wang, 1967). In adverse conditions, pitch serves as a major perceptual cue for separating competing speech streams and improving speech recognition in the presence of background talkers (Assmann, 1999; Bird & Darwin, 1998; Brokx & Nooteboom, 1982; Summers & Leek, 1998; Zekveld, Rudner, Kramer, Lyzenga, & Rönnberg, 2014).

Natural speech has variations in pitch (i.e., intonation), referred to here as *dynamic pitch*. As one of the prosodic cues, dynamic pitch plays an important role in facilitating speech comprehension in quiet (e.g., Brown, Salverda, Dilley, & Tanenhaus, 2011; Cutler, 1976; Steinhauer, Alter, & Friederici, 1999; Wingfield, Lombardi, & Sokol, 1984) as well as communicating emotion (Fairbanks, 1940; Frick, 1985). A number of recent studies have also found that natural pitch contour in speech facilitates speech recognition in noisy environments (e.g., Binns & Culling, 2007; Laures & Bunton, 2003; Miller, Schlauch, & Watson, 2010).

We know real listening environments are often noisy (Hodgson, Steininger, & Razavi, 2007; Olsen, 1998; Smeds, Wolters, & Rung, 2015). Recognizing speech in noise is a difficult task for many listeners, particularly those who are older and have hearing loss. Although dynamic pitch has the potential to help these listeners perceive speech better in noise, we still know little about whether older listeners with hearing loss can benefit from dynamic pitch in noise and how the benefit interacts with noise characteristics.

Dynamic Pitch and Speech Recognition in Temporally Modulated Noise

Dynamic pitch in target speech has consistently been shown to facilitate younger listeners' speech recognition in background noise (Assmann, 1999; Binns & Culling, 2007; Laures & Bunton, 2003; Laures & Weismer, 1999; Miller et al., 2010; Watson & Schlauch, 2008). For example, Laures and Bunton (2003) examined speech intelligibility with natural and flattened pitch contours using speech materials recorded from two male and two female speakers. The background noise was white noise in the first experiment and 12-talker babble in the second. Speech intelligibility scores were lower in the flattened-f0 condition than in

[a]Northwestern University, Evanston, IL

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

Editor: Julie Liss
Associate Editor: Mitchell Sommers

Received October 6, 2016
Revision received January 29, 2017
Accepted March 29, 2017
https://doi.org/doi.org/10.1044/2017_JSLHR-H-16-0389

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

the unmodified condition in both noise scenarios. Across the two noises, the amount of intelligibility reduction was equal for female speakers but greater with multitalker babble than white noise for male speakers.

In a series of experiments, Binns and Culling (2007) tested the effect of dynamic pitch on speech recognition in noise by creating multiple levels of dynamic-pitch strength ranging from the original pitch contour to the inverse pitch contour. Their findings suggest that reducing the amount of dynamic pitch in target speech has a progressively detrimental effect on speech intelligibility and that inverted dynamic pitch had the strongest negative impact. The effect was found to be stronger with a single-talker competing-speech masker than speech-shaped noise. In a more recent study, the detrimental effect of inverted pitch contour on speech intelligibility in noise was replicated by Miller et al. (2010). They also compared speech intelligibility with pitch manipulations of inverted pitch contour versus frequency-modulated f0. The finding was that the two pitch manipulations had a comparable effect and significantly reduced younger listeners' speech-recognition performance. This result suggests that the lack of linguistically correct dynamic-pitch cues contributes to reduced speech intelligibility in noise. Those researchers also found, interestingly, that exaggerating (by a factor of 1.75) and flattening pitch contours produced similar detrimental effects on speech intelligibility (although less than inverted pitch).

Although the effect of dynamic pitch on younger listeners' speech recognition in noise has been consistently demonstrated across multiple data sets, these studies either used only one type of noise (Assmann, 1999; Miller et al., 2010; Watson & Schlauch, 2008) or compared a temporally unmodulated noise with a speech masker (Binns & Culling, 2007; Laures & Bunton, 2003). It should be noted that speech maskers are different from speech-shaped noise not only in terms of temporal characteristics (i.e., modulated versus unmodulated) but also in terms of additional informational masking (e.g., Brungart, Simpson, Ericson, & Scott, 2001; Mattys, Brooks, & Cooke, 2009). Therefore, a comparison between a speech masker and an unmodulated noise is likely to include these two factors in explaining the benefit from dynamic-pitch cues.

We know that younger listeners with normal hearing can benefit from the modulation in fluctuating noise to recognize target speech (Festen & Plomp, 1990; Gustafsson & Arlinger, 1994; Lorenzi, Husson, Ardoint, & Debruille, 2006; Miller & Licklider, 1950). Temporal dips in fluctuating noise presumably provide moments that have favorable local signal-to-noise ratios (SNRs), and younger listeners with normal hearing can use these segments of target signal to process and recognize speech. This ability has been termed *glimpsing* (e.g., Cooke, 2006). Following this rationale, we hypothesized that temporal modulation in noise facilitates the use of dynamic pitch and provides an extra benefit for speech recognition. Experiment 1 tested this hypothesis while controlling for the informational masking in noise by using nonspeech noises with different temporal-modulation characteristics.

Dynamic Pitch and Older Listeners' Speech Recognition in Noise

Older adults' difficulty with understanding speech in noise is well documented, and this problem is likely due to multiple factors including peripheral, perceptual, and cognitive abilities (e.g., Dubno, Dirks, & Morgan, 1984; Helfer & Freyman, 2008; Humes, 1996; Pichora-Fuller & Souza, 2003; Plomp & Mimpen, 1979). We know that older listeners rely heavily on prosodic cues for speech comprehension in quiet (for a review, see Wingfield & Tun, 2001). Because these cues can potentially provide extra benefit when the listening environment becomes adverse, it is predicted that older listeners benefit from these cues for recognizing speech better in noise.

Many older listeners, particularly those with hearing loss, receive limited benefit from temporal modulation in noise (Dubno, Horwitz, & Ahlstrom, 2003; George, Festen, & Houtgast, 2006; Takahashi & Bacon, 1992; Wilson et al., 2010) and have difficulty understanding speech in these scenarios. This problem has been attributed to factors such as reduced high-frequency audibility (e.g., Bacon, Opie, & Montoya, 1998), poor temporal resolution that diminishes opportunities for glimpsing (e.g., Festen & Plomp, 1990; Gordon-Salant, 2005), and a more favorable baseline SNR that limits the glimpsing benefit (e.g., Bernstein & Grant, 2009). For instance, George et al. (2006) measured speech recognition in unmodulated and modulated background noises by listeners with mild–moderate hearing loss. They found that these listeners recognized speech better in modulated than unmodulated noise. The benefit from noise modulation, however, was substantially less than for listeners without hearing loss (or with simulated threshold elevation). Similar patterns have also been reported in older listeners with normal low-frequency hearing thresholds (Takahashi & Bacon, 1992).

Building on this literature, we hypothesized that if audible speech is a critical factor for benefiting from dynamic-pitch cues, the benefit from dynamic pitch would be stronger in modulated than unmodulated noise. Following this rationale, we would expect both speech-recognition performance and the benefit from dynamic pitch to vary with noise modulation for older listeners, albeit to a lesser extent than for younger listeners.

Further, although younger listeners do not benefit from a stronger dynamic pitch for speech recognition in unmodulated noise (see Miller et al., 2010), it could be beneficial for older listeners on three counts. First, because older listeners have poorer speech-in-noise performance in general, any extra perceptual cues that may be redundant for younger listeners could make a difference for the older listeners to process speech better. Second, there is evidence showing that older listeners use prosody more than younger listeners for sentence comprehension in quiet (Wingfield, Wayland, & Stine, 1992). Therefore, enhanced pitch may be particularly helpful for older listeners to process language and recognize speech. Last, data from listeners with profound hearing loss suggest that natural variations in f0 may be too small for them to perceive, and therefore a stronger intonation may be beneficial (Grant, 1987). Because many older listeners have significant hearing loss, it is possible that they would find a stronger pitch contour helpful.

Another aim of this experiment was to evaluate the impact of hearing loss and aging on older individuals' dynamic-pitch benefit. We know that older individuals' hearing ability negatively affects how well they can hear speech in fluctuating noise (e.g., Gordon-Salant, 2005) and whether pitch cues in speech can be perceived (for a review, see Moore & Carlyon, 2005). Older listeners' age has similarly been shown to be associated with speech-in-noise ability (e.g., Dubno et al., 1984) and temporal processing ability, which may contribute to pitch-perception deficit (e.g., Moore, 2008). Therefore, it was predicted that more hearing loss and more advanced age for the older listeners would be associated with less benefit from dynamic pitch, particularly in modulated noise.

Experiment 2 was carried out to examine the aforementioned three questions. First, it measured the effect of temporal modulation in noise on older listeners' benefit from dynamic-pitch cues for speech recognition in noise. Further, we used a stronger dynamic-pitch condition to test whether older listeners could benefit from this manipulation. In addition, this experiment included older participants who had a wide range of age and hearing status to evaluate the potential influence of these factors on dynamic-pitch benefit.

Dynamic-Pitch Perception and Speech-in-Noise Benefit

Peripheral hearing as measured by pure-tone average (PTA) is included as a measure of an individual's hearing ability in the present study, but it is also worth noting that this measure may not capture the listener's ability to perceive dynamic-pitch cues. To be specific, older listeners' dynamic-pitch perception could be degraded by various suprathreshold hearing deficits, which may include poor temporal coding (e.g., Grose & Mamo, 2010; Hopkins & Moore, 2011), degraded frequency selectivity (e.g., Matschke, 1991; Peters & Moore, 1992), and degraded neural representation of frequency modulation (e.g., Clinard & Cotter, 2015).

Indeed, our previous studies have demonstrated a large amount of individual variability in dynamic-pitch perception among older listeners with good hearing (Shen, Wright, & Souza, 2016; Souza, Arehart, Miller, & Muralimanohar, 2011). Although this variability may largely stem from suprathreshold and perceptual deficits of some older individuals with near-normal thresholds, we expect any deficits to be more prevalent in listeners with mild–moderate hearing loss (Buss, Hall, & Grose, 2004; Hopkins & Moore, 2011; Moore & Peters, 1992). In the present study, we hypothesized that individual ability to perceive dynamic-pitch contour would contribute to the benefit from dynamic-pitch cues for speech-in-noise performance. Experiment 3 examined the connection between individual dynamic-pitch perception (as measured by an intonation-identification task) and speech-in-noise benefit from these cues.

Experiment 1

Method

Participants

Eighteen younger adults (15 women, three men) aged 18–32 years (mean age = 22.6 years) participated in this study. All the listeners had normal hearing in both ears as defined by pure-tone thresholds ≤ 20 dB HL (American National Standards Institute, 2004) at all octave frequencies between 250 and 8000 Hz. The participants were recruited via flyers around the Northwestern University campus. All participants were native speakers of English with no experience in tonal languages and had no or minimal musical experience (less than 2 years of instrumental or vocal training).

The study consisted of a single 2-hr session, and the participants were compensated for their time. All participants completed an informed-consent process, and the institutional review board at Northwestern University approved the study protocol.

Stimuli

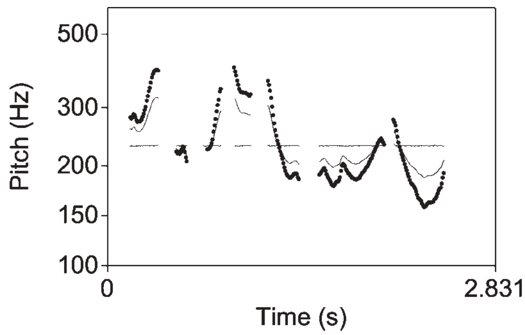
The target speech stimuli were drawn from Harvard IEEE sentences (Rothauser et al., 1969) that were produced by a female talker. These sentences have low predictability and therefore were ideal for minimizing the use of linguistic context for speech recognition. The dynamic-pitch contours of the stimuli were manipulated, and the sentences were resynthesized using the Praat program (Boersma & Weenink, 2013) with the method of pitch-synchronous overlap-and-add (Moulines & Charpentier, 1990). The purpose of this manipulation is to change the dynamic-pitch (f0) contour of the sentence while keeping other prosodic cues constant (e.g., duration, intensity). For each IEEE sentence, three f0 conditions were created using the following formula:

$$\begin{aligned}\text{Instant f0} &= \text{Sentence average f0}\\ &\quad + (\text{Original instant f0} - \text{Sentence average f0})\\ &\quad \times \text{Pitch factor.} \qquad (1)\end{aligned}$$

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 (see Miller et al., 2010). Figure 1 presents the f0 trajectories of the three dynamic-pitch conditions.

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. These nonspeech noises were created by filtering speech by a

Figure 1. Fundamental-frequency trajectories of the speech stimuli with the three dynamic-pitch levels (monotone: thin flat line; original: thin contour; strong: dotted contour).



three-band filter and randomly sign-reversing 50% of the signal in each band, which renders the output signal unintelligible while maintaining the original temporal-modulation property. The noises were then spectrally shaped by using the long-term spectrum of each talker condition (i.e., one-talker, two-talker). As shown by Dreschler et al., the one- and two-talker noises were very similar in terms of the long-term spectrum (with a difference of only approximately 2–3 dB in the frequency range of 100–200 Hz).

Procedure

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). The audiometric testing was done using an AC40 Interacoustics audiometer (Eden Prairie, MN) connected to ER-3 insert earphones (Elk Grove Village, IL).

Speech reception thresholds (SRTs) were obtained for all nine conditions (3 dynamic-pitch conditions × 3 background-noise conditions) using a customized MATLAB program (The MathWorks, Natick, MA) with an adaptive procedure (Plomp & Mimpen, 1979). The initial SNR increment was 6 dB until at least three out of five key words were repeated correctly. For each subsequent sentence, the SNR increased by 2 dB when zero to two key words were correctly repeated or decreased by the same amount for three to five correct key words. The number of trials was fixed at 15. The SNR varied across trials to track a performance level of 50% correct. SRTs for each condition were measured twice. When any testing run provided fewer than three reversals, or when the standard deviation across the final reversals exceeded 4 dB, SRT was measured a third time. Thresholds for each run were computed by taking the mean SNR (dB) across the reversals at the final step size of 2 dB. The order of conditions was counterbalanced across participants following a Latin-square design. Prior to testing, participants were given brief training on the different conditions to familiarize them with the different types of speech and noise. Practice consisted of nine trials and started at 0 dB SNR.

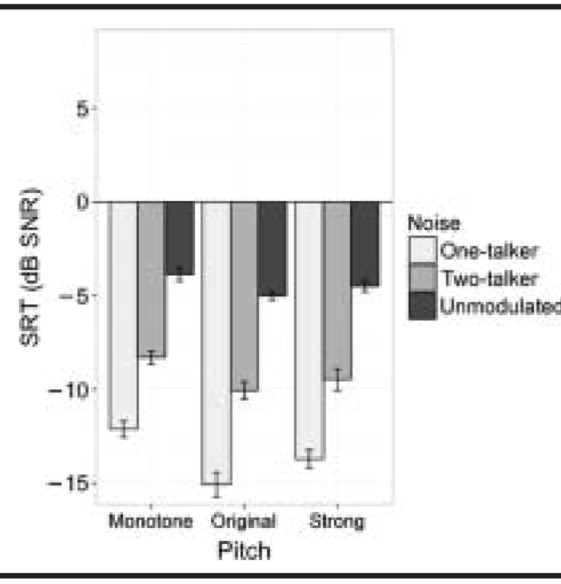
Stimuli were presented monaurally in the better ear at 68 dB SPL using a FastTrackPro external sound card (M-Audio, Cumberland, RI) and an ER-2 insert earphone (Etymotic Research, Elk Grove, IL) in the test ear. Listeners were seated in a double-walled soundproof booth. They were instructed to repeat back the sentences aloud for the experimenter, who scored the responses from outside the booth.

Results

The SRTs of the 18 younger listeners are presented in Figure 2. Data analysis was conducted using mixed-effects linear regression (Baayen, 2008) using R (Version 3.2.1) with the lme4 and lmerTest libraries (Bates, Maechler, Bolker, & Walker, 2014; Kuznetsova, Brockhoff, & Christensen, 2013).

The first model included fixed effects for dynamic pitch, noise modulation, and order of measurement, and random intercepts for participant and item. The dependent variable was SRT; order of measurement was included as a control variable and was contrast coded (−.5, .5). Helmert coding, a strategy for coding categorical variables in mixed-effects modeling, is used to compare the dependent variable on a specific level of an independent variable with the means of the dependent variable across the other two levels of the independent variable. In this model, dynamic pitch was Helmert coded: The first contrast compared differences between the monotone and the joint mean of the original-pitch and strong-pitch conditions (−.5, .25, .25); the second contrast compared differences between the original-pitch and strong-pitch conditions (0, .5, −.5). Noise modulation was also Helmert coded: The first contrast compared differences between the unmodulated noise and

Figure 2. Speech reception thresholds (SRTs) of the younger group (n = 18). Error bars indicate ±1 standard error. SNR = signal-to-noise ratio.



the joint mean of one- and two-talker noise conditions (−.5, .25, .25); the second contrast compared differences between the one- and two-talker conditions (0, .5, −.5). Table 1 reports chi-square and *p* values on the basis of the likelihood-ratio tests between the models with and without the predictor of interest. Both dynamic pitch and noise modulation (but not the interaction) significantly improved the model.

The effects of dynamic pitch and noise were further tested as fixed factors in the linear mixed model using *t* tests with Satterthwaite approximations to degrees of freedom (Kuznetsova et al., 2013). The results showed that the presence of dynamic pitch significantly improved speech recognition, $b = -2.04$, $SE = 0.32$, $t = -6.38$, $p < .001$, and that the original dynamic pitch was better than strong dynamic pitch, $b = -0.87$, $SE = 0.27$, $t = -3.23$, $p < .01$. The younger listeners performed significantly better in modulated noise than in unmodulated noise, $b = -9.08$, $SE = 0.32$, $t = -28.58$, $p < .001$, and in one-talker noise than in two-talker noise, $b = 4.39$, $SE = 0.27$, $t = 16.26$, $p < .001$.

In order to examine the efficacy of dynamic-pitch cues, further analysis was conducted regarding the amount of SRT benefit from dynamic-pitch cues (as defined by the SRT differences between the monotone condition and each of the two dynamic-pitch conditions). Figure 3 illustrates the SRT benefit scores for the younger group. The model included fixed effects for dynamic pitch, noise modulation, and SRT in the monotone condition and random by-participant intercepts. The SRT in the monotone condition served as a baseline score and could potentially influence the amount of benefit the listeners get from the dynamic-pitch cues (see Bernstein & Grant, 2009). Therefore, this continuous variable was included as a control variable. Dynamic pitch was contrast coded (−.5, .5). Noise modulation was Helmert coded: The first contrast compared differences between the unmodulated noise and the joint mean of the one- and two-talker noise conditions (−.5, .25, .25); the second contrast compared differences between the one- and two-talker conditions (0, .5, −.5). Table 1 reports chi-square and *p* values on the basis of the likelihood-ratio tests between the models with and without the predictor of interest. Both dynamic pitch and noise modulation significantly improved the model for predicting the amount of benefit from dynamic pitch, whereas the interaction did not.

The amount of benefit was larger for the original-pitch condition than for the strong-pitch condition, $b = -0.82$, $SE = 0.31$, $t = -2.59$, $p = .01$. The listeners benefited more from the presence of dynamic-pitch cues in modulated noise than in unmodulated noise, $b = 5.07$, $SE = 1.06$, $t = 4.77$, $p < .001$, and in one-talker noise than in two-talker noise, $b = 2.50$, $SE = 0.58$, $t = 4.29$, $p < .001$.

Discussion

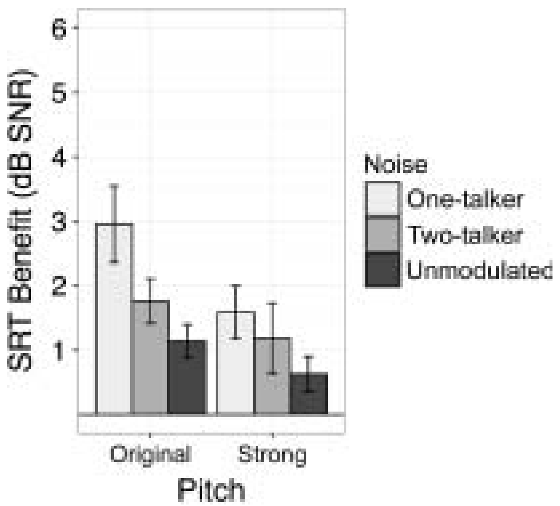
To our knowledge, the present study was the first to control the characteristics of the background noise in order to specifically test the hypothesis that temporal modulation in noise facilitates the use of dynamic pitch as a perceptual cue for speech recognition. Results from younger listeners with normal hearing support this hypothesis by showing that the benefit from dynamic-pitch cues varied with the amount of temporal modulation in noise and was significantly different between unmodulated and modulated noise as well as between noises with different modulation depths. This finding can be explained by the glimpsing account of speech recognition (e.g., Festen & Plomp, 1990; Miller & Licklider, 1950). When segments of speech are audible through temporary dips in modulated noises, younger listeners with normal hearing are able to exploit dynamic pitch to a larger extent to facilitate speech recognition. We also found that listeners benefit more with a strong pitch contour in modulated noise (i.e., with about 1.5 dB SNR benefit) but not unmodulated noise (i.e., with about 0.5 dB SNR benefit). First, this finding aligns with the results of Miller et al. (2010), in which the listeners performed at approximately the same level with strong and monotone pitch in unmodulated noise. Second, this result extends the previous literature by showing a small glimpsing effect that modulates the benefit from a strong pitch contour. Taken together with the previous finding that both inverted and frequency-modulated pitch contours are detrimental for recognizing speech in unmodulated noise (see Miller et al., 2010), the role of dynamic pitch in facilitating glimpsing is likely to stem from the positive effect of prosody on processing of the linguistic content.

Our findings further show that a strong dynamic-pitch contour had a detrimental effect on speech recognition as compared with the original pitch contour, particularly when the noise was strongly modulated. Overall, this result suggests that whereas the presence of dynamic-pitch cues facilitates speech recognition in noise, exaggerated dynamic pitch is likely to pose a detrimental effect, which could be due to unnaturalness in speech prosody and/or

Table 1. Model-comparison statistics of pitch and noise effects on speech reception thresholds and dynamic-pitch benefit scores for the younger group.

	Speech reception threshold			Benefit score		
Variable	χ^2	*df*	*p*	χ^2	*df*	*p*
Dynamic pitch	48.06	2	< .001	8.27	2	< .050
Noise modulation	418.00	2	< .001	22.29	2	< .001
Dynamic Pitch × Noise Modulation	6.70	4	> .100	1.59	4	> .100

Figure 3. Dynamic-pitch benefit (i.e., speech reception thresholds [SRTs] in the monotone condition minus SRTs in the original and strong conditions) of the younger group. Error bars indicate ±1 standard error. SNR = signal-to-noise ratio.



signal distortion from the synthesis process (which is discussed further in the next section). In addition, this detriment is particularly apparent when strong temporal modulation in noise (i.e., one-talker noise) makes dynamic-pitch cues (as well as the speech signal) highly accessible.

In order to find an approach that can minimize the detrimental effect of a stronger dynamic pitch, two potential sources of this effect have to be considered. First, our pitch manipulation on a macro scale does not faithfully represent the stronger dynamic pitch in natural speech production. For example, some pitch variations are associated with the linguistic meaning of the speech, such as pitch accent and intonation (e.g., Bolinger, 1958; Ladd & Cutler, 1983). Although these dynamic-pitch cues bear more importance than others, our manipulation exaggerates all pitch cues indiscriminately, which could potentially introduce unnaturalness in the processed speech with strong dynamic-pitch contours. Future work should explore the possibility of modeling naturally produced dynamic-pitch contours, which can preserve the natural prosody, and applying them in the speech-in-noise scenario.

As to the possibility of signal distortion, Miller et al. (2010) proposed the explanation that when the synthesized f0 is considerably higher than the original f0, the formant peaks of the vowels are reduced due to misalignment of harmonics of f0 with vocal-tract resonances (de Cheveigné & Kawahara, 1999; Diehl, Lindblom, Hoemeke, & Fahey, 1996). This distortion could lower speech intelligibility, particularly in background noise. It should be noted that this distortion is specifically associated with the extreme pitch values that raise the pitch level considerably (i.e., in the strong dynamic-pitch manipulation). Although this signal distortion could potentially contribute to the speech-recognition performance in the strong dynamic-pitch condition, it cannot account for our results across all three pitch conditions (speech-recognition performance was better with strong dynamic pitch than with monotonous speech even though dynamic pitch was manipulated in both conditions). Because the present study and that of Miller et al. (2010) both used a female talker, it is also possible that a high original f0 level has worsened the detrimental distortion. It is worth noting that Binns and Culling (2007) found that decreasing the pitch contour by 50% did not negatively affect speech intelligibility, whereas a decrease of 75% did. This is at least partially due to the deprivation of dynamic-pitch cues, but another factor that could play a role here is that the unfavorable distortion is a function of the amount of pitch change. This would lend itself to the empirical question whether increasing the pitch contour by an amount that is less than 75% could induce a different effect. Further research should investigate these questions by using male voices and multiple values of stronger dynamic pitch in speech-intelligibility testing.

Whereas both the study by Miller et al. (2010) and the present study had younger participants with normal hearing, it has yet to be investigated whether older listeners are able to benefit from dynamic-pitch cues and whether the benefit varies depending on the temporal modulation in noise. These questions are examined in Experiment 2.

Experiment 2

Method

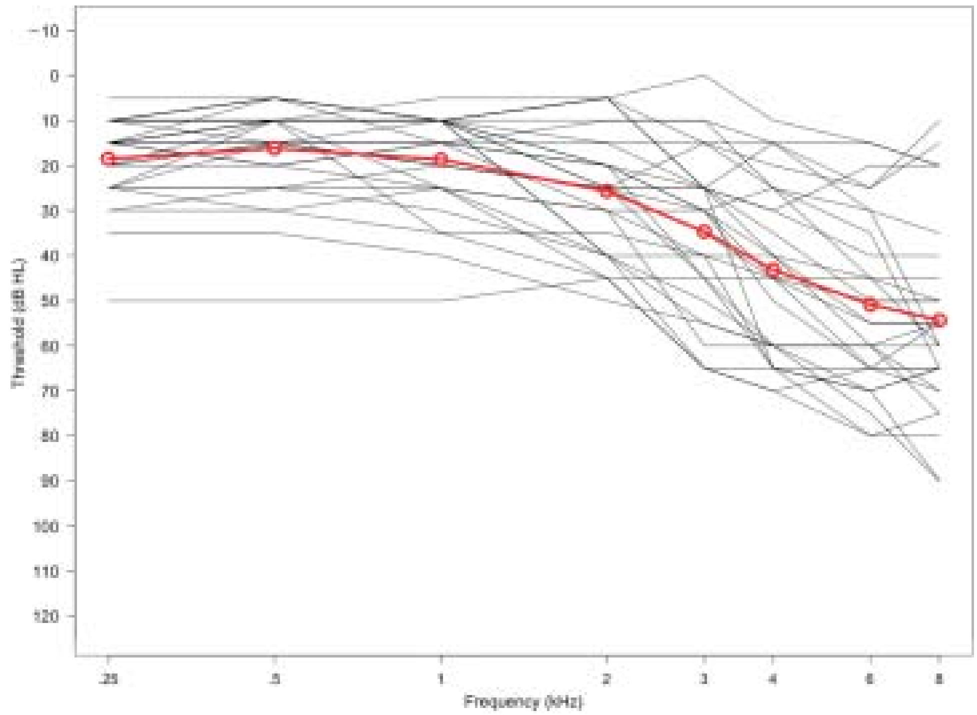
Participants

Thirty older adults (15 women, 15 men) aged 56–83 years (mean age = 70.2 years) participated in this study. Their hearing status ranged from near-normal hearing to mild–moderate sensorineural hearing loss (see Figure 4 for audiograms). Inclusion criteria included symmetrical hearing as defined by a difference in PTA (0.5, 1, 2 kHz) ≤ 10 dB across ears and air–bone gaps ≤ 10 dB at octave frequencies between 0.5 and 3 kHz.

The participants were recruited via newspaper advertisement and flyers in the Greater Chicago area. All participants were native speakers of English with no experience in tonal languages and had no or minimal musical experience (less than 2 years of instrumental or vocal training). Older participants were screened for mild cognitive impairment using the Montreal Cognitive Assessment (Nasreddine et al., 2005). All participants passed the test with a cutoff score of 23, which has been shown to maximize the test's sensitivity and specificity (Lee et al., 2008; Luis, Keegan, & Mullan, 2009).

The study setup and protocol were identical to Experiment 1. To accommodate each individual's hearing thresholds, we amplified the stimuli using the National Acoustics Laboratories–Revised linear prescriptive formula on the basis of individual thresholds (Byrne, Dillon, Ching, Katsch, & Keidser, 2001).

Figure 4. Audiograms of the older listeners (n = 30). Individual thresholds: thin lines; group average: red line with circles.



Results

The 30 older listeners' SRT data are plotted in Figure 5. A linear mixed-effects model was built using the same structure as in Experiment 1. Because age and PTA strongly covaried in the present data set (as indicated by a significant correlation between the two factors, $r = .513$, $p < .01$), the model was built with only one of these two fixed factors included at a time. Table 2 presents the likelihood-ratio test results between the models with and without the predictor of interest. Consistent with the younger listeners' data, both dynamic pitch and noise modulation were significant predictors of the SRT, and the interaction was not. PTA was found to be a significant predictor of speech-recognition performance.

As to the effects of dynamic pitch and noise modulation, we found the same pattern as in Experiment 1. Dynamic-pitch cues improved older listeners' speech recognition, $b = -3.72$, $SE = 0.31$, $t = -12.12$, $p < .001$, and performance was significantly better with the original dynamic pitch than the strong dynamic pitch, $b = -1.15$, $SE = 0.26$, $t = -4.36$, $p < .001$. Temporal modulation in noise was found to be beneficial for older listeners' speech recognition, $b = -4.29$, $SE = 0.31$, $t = -14.07$, $p < .001$, and performance was better with more modulation, $b = 3.07$, $SE = 0.26$, $t = 11.67$, $p < .001$. It is unsurprising that significant influences on overall performance of speech recognition in noise were found for older listeners' PTA (without inclusion of age), $b = 0.44$, $SE = 0.07$, $t = 5.98$, $p < .001$, and age (without inclusion of PTA), $b = 0.38$, $SE = 0.13$, $t = 3.02$, $p < .01$.

The older listeners' dynamic-pitch benefit data are presented in Figure 6. Taking dynamic-pitch benefit score

Figure 5. Speech reception thresholds (SRTs) of the older group (n = 30). Error bars indicate ±1 standard error. SNR = signal-to-noise ratio.

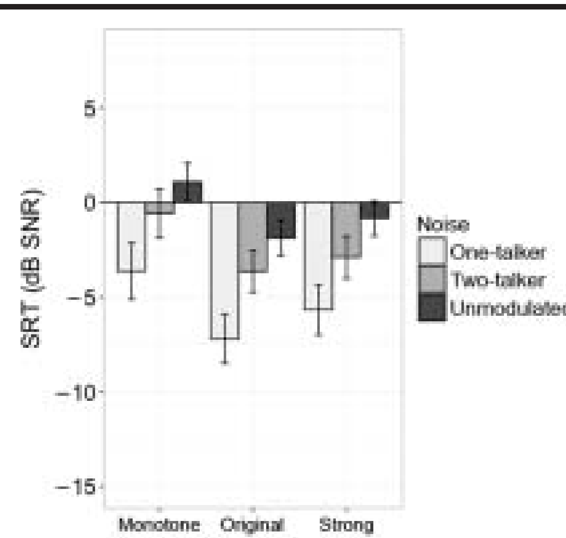


Table 2. Model-comparison statistics of pitch and noise effects on speech reception thresholds and dynamic-pitch benefit scores for the older group.

	Speech reception threshold			Benefit score		
Variable	χ^2	*df*	*p*	χ^2	*df*	*p*
Dynamic pitch	143.40	2	< .001	13.84	2	< .01
Noise modulation	113.87	2	< .001	12.56	2	< .05
Dynamic Pitch × Noise Modulation	1.64	4	> .100	4.65	4	> .10
PTA	24.68	1	< .001	3.55	1	> .05
Age (without PTA in the model)	8.42	1	< .010	0.86	1	> .05

Note. PTA = pure-tone average.

as the dependent variable, the model included random by-participant intercepts and the following fixed effects: dynamic pitch, noise modulation, SRT in the monotone condition, and PTA (another model was built with age instead of PTA). Table 2 presents the likelihood-ratio test results between the models with and without the predictor of interest. Neither PTA ($\chi^2 = 3.55$, $p = .059$) nor age ($\chi^2 = 0.86$, $p > .1$) was found to be a significant factor in predicting dynamic-pitch benefit.

The dynamic-pitch benefit model showed that the benefit from original dynamic pitch was larger than that from strong dynamic pitch, $b = -1.10$, $SE = 0.30$, $t = -3.62$, $p < .001$. The dynamic-pitch benefit was lower in unmodulated noise than in modulated noise, $b = 1.45$, $SE = 0.46$, $t = 3.15$, $p < .01$, and in two-talker noise than in one-talker noise, $b = 0.86$, $SE = 0.39$, $t = 2.21$, $p < .05$.

Because our older participants had varying degrees of hearing sensitivity, ranging from normal to moderate hearing loss, it is likely that the impact of hearing loss on dynamic-pitch benefit was stronger for the subgroup of older listeners with significant hearing loss. These listeners are at a disadvantage in terms of speech recognition in noise, so it is of particular importance to know whether and how much they can benefit from dynamic pitch. For this reason, we built another set of models (using the same structure as those in Experiment 1) on SRT (see Figure 7) and dynamic-pitch benefit scores (see Figure 8). Data included in this set of models were from 14 older listeners with hearing loss, as defined by a pure-tone threshold > 25 dB at any interoctave frequency between 0.25 and 2 kHz and/or > 35 dB at 3 kHz.

The SRT model included fixed effects for dynamic pitch, noise modulation, and order of measurement, and random by-participant and by-item intercepts. Both pitch contrasts were significant (monotone vs. dynamic pitch: $b = -4.55$, $SE = 0.48$, $t = -9.51$, $p < .001$; original vs. strong pitch: $b = -1.09$, $SE = 0.41$, $t = -2.64$, $p < .01$), and so were the noise-modulation contrasts (unmodulated vs. modulated: $b = -2.88$, $SE = 0.48$, $t = -6.0$, $p < .001$;

Figure 6. Dynamic-pitch benefit (i.e., speech reception thresholds [SRTs] in the monotone condition minus SRTs in the original and strong conditions) of the older group. Error bars indicate ±1 standard error. SNR = signal-to-noise ratio.

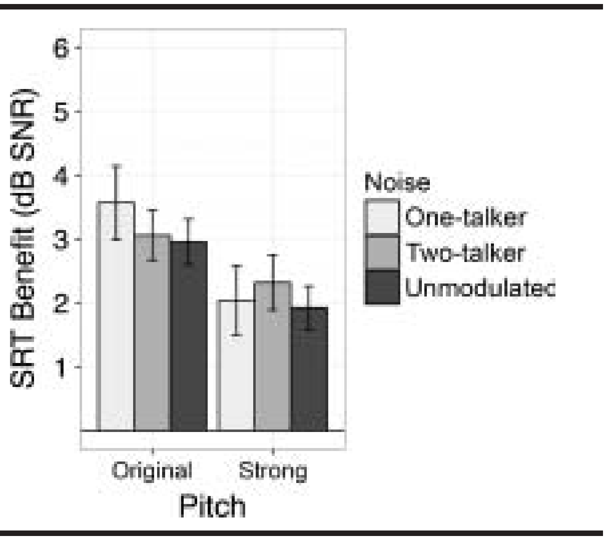


Figure 7. Speech reception thresholds (SRTs) of the older group with hearing loss (n = 14). Error bars indicate ±1 standard error. SNR = signal-to-noise ratio.

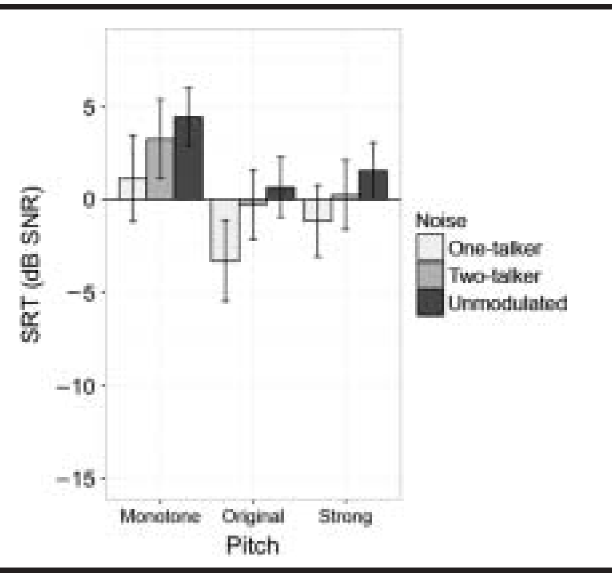
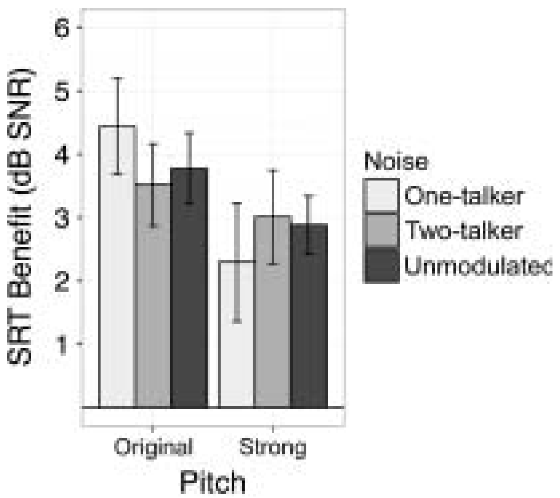


Figure 8. Dynamic-pitch benefit (i.e., speech reception thresholds [SRTs] in the monotone condition minus SRTs in the original and strong conditions) of the older group with hearing loss. Error bars indicate ±1 standard error. SNR = signal-to-noise ratio.



one-talker vs. two-talker: $b = 2.26$, $SE = 0.41$, $t = 5.43$, $p < .001$). None of the interactions between pitch strength and noise modulation were significant.

The benefit-score model was built with random by-participant intercepts and the following fixed effects: dynamic pitch, noise modulation, PTA, and SRT in the monotone condition (as a control variable). The results revealed a strong effect of PTA, $b = -0.23$, $SE = 0.07$, $t = -3.33$, $p < .01$, in the direction of more hearing loss being associated with less dynamic-pitch benefit. Noise modulation, however, did not have a significant impact on the amount of benefit for these older listeners (unmodulated vs. modulated: $b = 1.13$, $SE = 0.72$, $t = 1.57$, $p = .12$; one-talker vs. two-talker: $b = 0.92$, $SE = 0.61$, $t = 1.50$, $p = .14$).

In order to examine the potential interaction between noise modulation and group (i.e., older and younger listeners), we built another set of models with data from both younger and older groups. The SRT model included fixed effects for dynamic pitch, noise modulation, group (younger vs. older), Group × Noise interaction, and order of measurement, and random intercepts of participant and item. The benefit model had dynamic-pitch benefit as the dependent variable and included random by-participant intercepts and the following fixed effects: dynamic pitch, noise modulation, group, Group × Noise interaction, and SRT in the monotone condition. The model-comparison results are reported in Table 3. They show that besides the effects of dynamic pitch and noise modulation, which significantly improved both models, the Group × Noise modulation interaction was a significant predictor in both SRT and dynamic-pitch benefit models, suggesting a reduced effect of noise modulation on both SRTs and benefit scores in older as compared with younger listeners.

Discussion

To our knowledge the present study is the first one showing that older listeners with varying degrees of hearing sensitivity can benefit from dynamic pitch for speech recognition in noise. Although older listeners were able to benefit from the dynamic-pitch cues, their benefit was influenced by temporal modulation of noise to a lesser extent than for younger listeners. To be specific, older listeners as a group had less difference in dynamic-pitch benefit between modulated and unmodulated noise than did younger listeners (see Table 3). This finding will be addressed in more detail in the General Discussion.

Contrary to our prediction, older listeners as a group did not benefit from an exaggerated pitch contour more than the original contour in recognizing speech. It is worth noting that our data revealed large variability across older individuals in terms of benefit from stronger dynamic-pitch cues, particularly in the highly modulated noise. In other words, whereas some older listeners appeared to benefit more from the original pitch contour, others did better with the strong pitch contour (for a more detailed description of data from older listeners with hearing loss, see Shen & Souza, in press). This finding suggests that although the potential distortion associated with stronger pitch manipulation also negatively affected older listeners' speech intelligibility, some of them performed much better with this pitch exaggeration, to the extent that the benefit counteracted the detriment. Follow-up research is warranted to investigate the potential of individualized optimal dynamic-pitch strength.

For those older listeners with significant hearing loss, their hearing as defined by PTA was found to negatively associate with the dynamic-pitch benefit, after the baseline performance with speech in noise was controlled for. It is worth noting that we compensated for each individual's elevated threshold by amplifying the stimuli. Therefore, the effect of hearing loss on reduced benefit from dynamic pitch is likely to stem from the suprathreshold deficits that are commonly observed in individuals with hearing loss. It is possible that the degraded hearing abilities of our older listeners were detrimental to their pitch perception and therefore reduced the benefit from dynamic-pitch cues for speech recognition in noise. Experiment 3 examined this hypothesis by measuring older individuals' dynamic-pitch perception using an intonation-perception task.

Experiment 3

Method

Participants

The same group of 30 older participants as in Experiment 2 participated in this experiment. This experiment comprised one 2-hr session.

Stimuli

The stimuli consisted of two monophthongs (/ɑ/, /i/) and two diphthongs (/ɑi/, /iɑ/). These tokens were 620 ms

Table 3. Model-comparison statistics of pitch, noise, and group effects on speech reception thresholds and dynamic-pitch benefit scores for all participants.

Variable	Speech reception threshold			Benefit score		
	χ^2	*df*	*p*	χ^2	*df*	*p*
Dynamic pitch	195.81	2	< .001	18.13	2	< .001
Noise modulation	605.48	2	< .001	37.89	2	< .001
Group (older vs. younger)	126.06	1	< .001	7.95	1	.050
Group × Noise	146.10	3	< .001	7.86	3	< .050

long. They were modeled on a single male talker from the northern cities of the United States and synthesized using a speech synthesizer in cascade mode. The stimuli had an f0 in the range of 80–160 Hz, and the f0 at the stimulus midpoint in time was kept at 113 Hz. The ratio of starting f0 to ending f0 varied in six equal logarithmic steps from 1:0.5 to 1:2.

Procedure

We used a two-alternative forced-choice identification paradigm to measure the participants' ability to identify dynamic pitch with rising or falling glides. On each trial, a stimulus was presented, and the participant was asked to select the button that corresponded to the direction of pitch change (rise or fall). Each participant had a practice block with feedback prior to two testing blocks. A testing block consisted of 168 trials, and a practice block consisted of 24. The order of the stimuli was randomized across participants. No feedback was given in the testing blocks.

All signals were sent from a custom MATLAB program to a Tucker-Davis Technologies (Alachua, FL) digital signal processor for digital-to-analog conversion. The signals were then routed through a programmable attenuator before being delivered to an ER-2 insert earphone. The presentation level, which was determined by pilot testing, was 70 dB SPL for participants with normal hearing and 35 dB SL for participants with hearing loss (with a maximum output level of 85 dB SPL).

Results

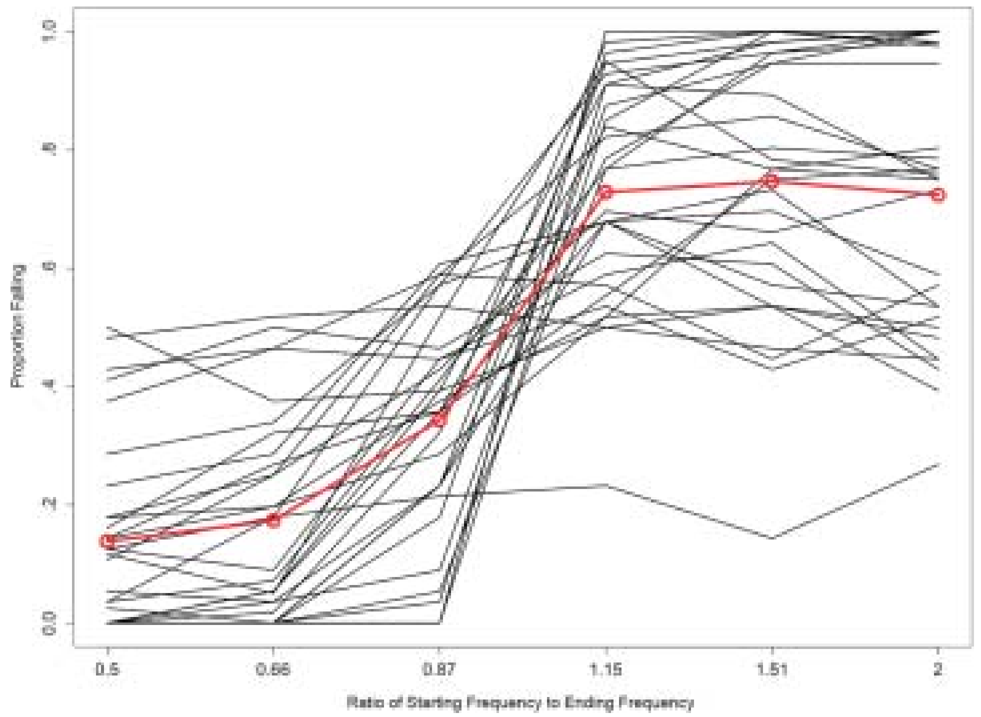
The dynamic-pitch perception data are presented in Figure 9 as individual psychometric functions with the proportion of responses falling as a function of the dynamic-pitch conditions, as indicated by the ratio between starting f0 and ending f0. To quantify dynamic-pitch perception, the f0 ratio was log-transformed with base 10, and logistic regression was fitted on each participant's data. The slope of each individual's psychometric function was used for the analysis, where a higher value indicates better performance and a lower value indicates worse performance.

A linear mixed-effects model was built with the dynamic-pitch benefit as the dependent variable (as defined by the SRT differences between the monotone condition and each of the two dynamic-pitch conditions). The model included fixed effects for dynamic pitch, noise modulation, SRT score in the monotone condition, and psychometric-function slope for the intonation-identification test, and random by-participant intercepts. The contrasts were set using the same method as in Experiment 1. Table 4 reports the chi-square and *p* values on the basis of the likelihood-ratio tests between the models with and without the predictor of interest. The data show that individuals' ability to perceive dynamic pitch (as defined by the slope of the psychometric function of the intonation-identification performance) was not a significant predictor for the amount of benefit the older listeners had from the dynamic-pitch cues, $b = 0.01$, $SE = 0.01$, $t = 1.08$, $p > .05$.

Discussion

Experiment 3 examined the relationship between dynamic-pitch perception and the amount of benefit from dynamic-pitch cues for speech recognition. It was hypothesized that when listeners can better perceive dynamic pitch in speech, they should be able to benefit from these cues more in a speech-in-noise scenario. Using an intonation-identification task, we found that this measure of dynamic-pitch perception did not predict individuals' benefit from these cues for speech recognition in noise. This result can potentially be explained by two factors. First, as suggested by findings from previous studies (Binns & Culling, 2007; Miller et al., 2010), the benefit from dynamic-pitch cues for speech recognition in noise may come from its role in facilitating online language processing by helping the listener direct attention to the stressed words and parse sentences more efficiently (Cutler, 1976; Wingfield et al., 1984). This process may engage cognitive abilities that are critical for online language processing, whereas the perceptual task we used here captures only pitch perception without linguistic context. Second, whereas the dynamic-pitch cues were presented in noise in the speech-recognition task, the intonation-identification task measured only the ability to perceive these cues in quiet. This mismatch raises the question as to whether those older listeners who can perceive dynamic pitch in quiet still miss it in noise and therefore cannot benefit from this cue for speech recognition in noise. Third, the perceptual ability to actively discern pitch-change direction (i.e., rising and falling) in short segments of speech is different from the ability to perceive dynamic-pitch cues in continuous speech and use them for speech comprehension. These possibilities should be investigated by future study.

Figure 9. Older listeners' psychometric functions for dynamic-pitch perception (individual functions: thin lines; group average: red line with circles).



General Discussion

The present study examined younger and older listeners' ability to recognize speech in noise with varying strengths of dynamic pitch in target speech and temporal modulation in nonspeech noise. Overall, the data were consistent with previous research (e.g., Binns & Culling, 2007; Laures & Bunton, 2003; Miller et al., 2010) in demonstrating the benefit from dynamic-pitch cues for speech recognition in noise. Data from younger and older listeners showed the same pattern: Speech-recognition performance was best with original dynamic pitch, followed by strong dynamic pitch, and worst with monotone speech. Although this result emphasizes the necessity for preserving dynamic-pitch cues, particularly for those older listeners who have difficulty hearing speech in noise, it is worth noting that the pitch manipulations tested in the present study (i.e., monotone, strong) are quite extreme and therefore unlikely to represent any natural signal distortions. Our study, however, serves as a first step in this line of work by testing these two extreme values (i.e., monotone and 1.75 times the original dynamic-pitch contour), which were chosen on the basis of findings from previous studies (Grant, 1987; Binns & Culling, 2007; Miller et al., 2010) and are likely to be boundary parameters on the continuum of dynamic-pitch strength. Although neither manipulation was found to be beneficial compared with the original dynamic pitch, there are potential implications that justify further testing of moderate dynamic-pitch values in this paradigm. It is unusual to encounter monotonous speech in reality, but speech that has reduced dynamic pitch is documented in individuals who have speech pathologies (e.g., dysarthria; Schlenck, Bettrich, & Willmes, 1993). A potential intervention for improving speech recognition with this type of speech is to strengthen dynamic pitch and bring it closer to natural strength. On the other side of the continuum, the strong dynamic-pitch manipulation in the present study inevitably rendered the speech unnatural, because the pitch value was extreme. It is still possible, however, that dynamic pitch that is stronger than a natural contour could be beneficial for speech recognition in noise, given that the value is relatively moderate. Further research is warranted to test these possibilities.

Consistent with our hypothesis, data from younger listeners showed that more temporal modulation in noise was associated with greater speech-recognition benefit from dynamic-pitch cues. This finding is in accordance

Table 4. Model-comparison statistics of pitch and noise effects on dynamic-pitch benefit scores for the older group.

	Benefit score		
Variable	χ^2	*df*	*p*
Dynamic pitch	13.33	2	< .01
Noise modulation	10.30	2	< .05
Dynamic Pitch × Noise Modulation	1.12	4	> .10
Dynamic-pitch perception	1.19	1	> .10

with previous data (Binns & Culling, 2007; Laures & Bunton, 2003) showing that dynamic-pitch benefit was more prominent with speech maskers than with speech-shaped unmodulated noise. It suggests that the additional benefit from dynamic pitch in a speech masker is at least partially due to the temporal modulation of the speech masker.

An interesting finding of our study was that the effect of noise modulation on dynamic-pitch benefit was reduced for the group of 30 older listeners. This finding can be explained by compromised temporal resolution, because if older listeners are not able to access the speech through dips in modulated noise, they will neither recognize speech better nor use dynamic-pitch cues more in modulated than in unmodulated noise. Our data support this rationale by showing a reduced noise-modulation effect on SRTs and dynamic-pitch benefit for the older group as compared with the younger group. This finding is consistent with the literature in suggesting older listeners' difficulty with speech perception in temporally modulated noise (e.g., Dubno et al., 2003; Festen & Plomp, 1990; Takahashi & Bacon, 1992).

When we focus on those 14 older listeners with significant hearing loss, the data show a different pattern. As indicated by the effect of noise modulation on SRT, these listeners were still able to glimpse, albeit with poorer ability than younger listeners. They were not, however, able to benefit more from dynamic pitch in modulated noise than unmodulated noise. In addition to the explanation that a decline in glimpsing ability reduces dynamic-pitch benefit in modulated noise, we think there are two other possible reasons that could account for this null effect of noise modulation on dynamic-pitch benefit in this group of older listeners with hearing loss. First, the older listeners with hearing loss could only perform the speech-recognition task with SNRs that were much more favorable than those for younger listeners. Under these SNRs, the effect of noise modulation on dynamic-pitch benefit may be less salient. However, note that this effect persists when individuals' baseline SNR levels are controlled for (by including individual baseline SRT in the model). Therefore, we think there may be another contributing factor, which is that some older listeners had poor abilities to use dynamic-pitch cues in modulated noise. This means that even though they can hear speech in the dips, they are less able to perceive and/or use dynamic pitch to understand speech better in fluctuating noise. According to this account, the older listeners would not have more benefit from dynamic-pitch cues in modulated noise than in unmodulated noise—which is shown by our data from older listeners with significant hearing loss. This result suggests that older listeners with hearing loss are not getting the benefit from dynamic-pitch cues even when the background noise provides access to these cues. We know speech is highly redundant in terms of acoustic cues, and a normal auditory/cognitive system is able to perceive and decode a speech signal even with a limited number of cues available. The focus of the present study was the role of dynamic pitch in speech perception, and we compared speech-recognition performance with and without this cue. Therefore, even if an older listener is able to use other cues (e.g., segmental cues) to recognize speech better in modulated noise, that listener could still have difficulty benefiting from the dynamic pitch, which is a cue that can improve speech-recognition performance in noise. Although the mechanism behind this phenomenon awaits further research, possible contributors include older listeners' hearing deficits as well as perceptual and cognitive difficulties.

One of our goals was to examine whether any individual factors can predict dynamic-pitch benefit in general. To this end, we found that for the older group with hearing loss, the amount of benefit from dynamic pitch was predicted by listener hearing (i.e., PTA). It should be noted that whereas hearing was measured by peripheral sensitivity here, this connection is likely to stem from the degraded suprathreshold functions that are usually observed in parallel with the reduced peripheral sensitivity. These suprathreshold hearing deficits can potentially make pitch cues in noise less accessible to older listeners with hearing loss. This result aligns with previous findings and suggests the effect of suprathreshold hearing abilities on older listeners' speech perception. For instance, it has been demonstrated that older listeners with speech-recognition difficulty have poorer spectral resolution of complex signals (Phillips, Gordon-Salant, Fitzgibbons, & Yeni-Komshian, 2000). Furthermore, Mackersie, Prida, and Stiles (2001) have found an association between speech-recognition performance in background speech and the ability to perceive a tone sequence as separate streams, which appeared to be difficult for listeners with hearing loss. Follow-up research is needed to investigate these possible connections.

Further, the ability to perceive dynamic-pitch cues (as measured by an intonation-identification task) was not a strong predictor for older individuals' benefit from dynamic pitch cues. This finding can potentially be explained by two factors. First, dynamic pitch is mostly helpful as a prosodic cue in a language context (Binns & Culling, 2007; Miller et al., 2010), and our intonation-identification task does not measure individuals' ability in a complex linguistic context. Second, our intonation-identification task presents dynamic-pitch stimuli without any background noise. This method does not capture an individual's ability to perceive it in noise, which may be more relevant to the amount of benefit for speech recognition in noise. These possibilities should be investigated by future research.

Last, the present data set showed that whereas age could predict older listeners' speech-in-noise performance, it was not associated with the amount of dynamic-pitch benefit that older individuals could get. Note that age and hearing were strongly correlated in the present data set, and therefore we could not directly evaluate any potential impact from individuals' cognitive and perceptual abilities without confounding the effect of hearing loss. Another possibility is that these abilities are not homogeneously associated with aging processes and may vary greatly across individuals. Further research is needed to shed light on whether aging processes affect the use of dynamic-pitch cues in perceiving speech with background noise.

Conclusions

The present study contributed to the literature by demonstrating that the effect of dynamic-pitch cues on speech recognition in noise depends on temporal modulation in background noise. Older listeners were able to benefit from the dynamic-pitch cues, but individuals varied substantially in terms of the dynamic-pitch strength they received the largest benefit from. For those older listeners with hearing loss, the amount of hearing loss strongly predicted the dynamic-pitch benefit for speech recognition in noise.

From a clinical perspective, these results provide evidence for a few implications and directions for future work. First, it is important to preserve dynamic-pitch cues in target speech, particularly in fluctuating background noise. Second, it is worth exploring the possibility of individualized optimal dynamic-pitch strengths to aid older listeners' speech perception in noise. Last, more effort should be devoted to helping older listeners with hearing loss use the dynamic-pitch cues in noise, perhaps starting with knowing more about how this ability is affected by hearing loss.

Acknowledgments

This work was supported by 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 study design; Shannon Switzer, Arleen Li, Laura Mathews, and Paul Reinhart for assistance with data collection; and Tim Schoof for help with the experimental program and comments on the manuscript. A portion of the data was presented at the 43rd Annual Scientific and Technology Conference of the American Auditory Society, Scottsdale, AZ, and the 2016 Hearing Across the Lifespan conference, Cernobbio, Italy.

References

American National Standards Institute. (2004). *Methods for manual pure-tone threshold audiometry* (ANSI S3.21-2004). New York, NY: Author.

Assmann, P. F. (1999). Fundamental frequency and the intelligibility of competing voices. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *14th International Congress of Phonetic Sciences* (pp. 179–182). Berkeley, CA: The Regents of the University of California.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, United Kingdom: Cambridge University Press.

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.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and S4 (Version 1.1-11) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/lme4/index.html

Bernstein, J. G. W., & 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,* 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,* 1765–1776.

Bird, J., & Darwin, C. J. (1998). Effects of a difference in fundamental frequency in separating two sentences. In A. R. Palmer, A. Rees, A. Q. Summerfield, & R. Meddis (Eds.), *Psychophysical and physiological advances in hearing* (pp. 263–269). London, United Kingdom: Whurr.

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/

Bolinger, D. L. (1958). A theory of pitch accent in English. *Word, 14,* 109–149.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics, 10,* 23–26.

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,* 1189–1196.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 110,* 2527–2538.

Buss, E., Hall, J. W., III, & Grose, J. H. (2004). Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss. *Ear and Hearing, 25,* 242–250.

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,* 37–51.

Clinard, C. G., & Cotter, C. M. (2015). Neural representation of dynamic frequency is degraded in older adults. *Hearing Research, 323,* 91–98.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America, 119,* 1562–1573.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics, 20,* 55–60.

de Cheveigné, A., & Kawahara, H. (1999). Missing-data model of vowel identification. *The Journal of the Acoustical Society of America, 105,* 3497–3508.

Diehl, R. L., Lindblom, B., Hoemeke, K. A., & Fahey, R. P. (1996). On explaining certain male–female differences in the phonetic realization of vowel categories. *Journal of Phonetics, 24,* 187–208.

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. *Audiology, 40,* 148–157.

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America, 76,* 87–96.

Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2003). Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing. *The Journal of the Acoustical Society of America, 113,* 2084–2094.

Fairbanks, G. (1940). Recent experimental investigations of vocal pitch in speech. *The Journal of the Acoustical Society of America, 11,* 457–466.

Faulkner, A., & Rosen, S. (1999). Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception. *The Journal of the Acoustical Society of America, 106,* 2063–2073.

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,* 1725–1736.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97,* 412–429.

George, E. L. J., 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,* 2295–2311.

Gordon-Salant, S. (2005). Hearing loss and aging: New research findings and clinical implications. *Journal of Rehabilitation Research & Development, 42*(4, Suppl. 2), 9–24.

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.

Grose, J. H., & Mamo, S. K. (2010). Processing of temporal fine structure as a function of age. *Ear and Hearing, 31,* 755–760.

Gustafsson, H. Å., & Arlinger, S. D. (1994). Masking of speech by amplitude-modulated noise. *The Journal of the Acoustical Society of America, 95,* 518–529.

Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing, 29,* 87–98.

Hodgson, M., Steininger, G., & Razavi, Z. (2007). Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments. *The Journal of the Acoustical Society of America, 121,* 2023–2033.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *The Journal of the Acoustical Society of America, 109,* 764–774.

Hopkins, K., & Moore, B. C. J. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *The Journal of the Acoustical Society of America, 130,* 334–349.

Humes, L. E. (1996). Speech understanding in the elderly. *Journal of the American Academy of Audiology, 7,* 161–167.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (Version 2.0-32) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/lmerTest/index.html

Ladd, D. R., & Cutler, A. (1983). Introduction: Models and measurements in the study of prosody. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 1–10). Berlin, Germany: Springer.

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,* 449–464.

Laures, J. S., & Weismer, G. (1999). The effects of a flattened fundamental frequency on intelligibility at the sentence level. *Journal of Speech, Language, and Hearing Research, 42,* 1148–1156.

Lee, J.-Y., Lee, D. W., Cho, S.-J., Na, D. L., Jeon, H. J., Kim, S.-K., . . . Cho, M. J. (2008). Brief screening for mild cognitive impairment in elderly outpatient clinic: Validation of the Korean version of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology, 21,* 104–110.

Lorenzi, C., Husson, M., Ardoint, M., & Debruille, X. (2006). Speech masking release in listeners with flat hearing loss: Effects of masker fluctuation rate on identification scores and phonetic feature reception. *International Journal of Audiology, 45,* 487–495.

Luis, C. A., Keegan, A. P., & Mullan, M. (2009). Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the Southeastern US. *International Journal of Geriatric Psychiatry, 24,* 197–201.

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,* 19–28.

Matschke, R. G. (1991). Frequency selectivity and psychoacoustic tuning curves in old age. *Acta Oto-Laryngologica, 111*(Suppl. 476), 114–119.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59,* 203–243.

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America, 22,* 167–173.

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,* 435–443.

Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology, 9*(4), 399–406.

Moore, B. C. J., & Carlyon, R. P. (2005). Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 234–277). New York, NY: Springer.

Moore, B. C. J., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *The Journal of the Acoustical Society of America, 91,* 2881–2893.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9,* 453–467.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53,* 695–699.

Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report. *American Journal of Audiology, 7*(2), 21–25.

Peters, R. W., & Moore, B. C. J. (1992). Auditory filter shapes at low center frequencies in younger and elderly hearing-impaired subjects. *The Journal of the Acoustical Society of America, 91,* 256–266.

Phillips, S. L., Gordon-Salant, S., Fitzgibbons, P. J., & Yeni-Komshian, G. (2000). Frequency and temporal resolution in elderly listeners with good and poor word recognition. *Journal of Speech, Language, and Hearing Research, 43,* 217–228.

Pichora-Fuller, M. K., & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology, 42*(Suppl. 2), 11–16.

Plomp, R., & Mimpen, 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,* 1333–1342.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., & Weinstock, M. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17,* 225–246.

Schlenck, K.-J., Bettrich, R., & Willmes, K. (1993). Aspects of disturbed prosody in dysarthria. *Clinical Linguistics & Phonetics, 7,* 119–128.

Shen, J., & Souza, P. (in press). Do older listeners with hearing loss benefit from dynamic pitch for speech recognition in noise? *American Journal of Audiology.* https://doi.org/10.1044/2017_AJA-16-0137

Shen, J., Wright, R., & Souza, P. E. (2016). On older listeners' ability to perceive dynamic pitch. *Journal of Speech, Language, and Hearing Research, 59*(3), 572–582.

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology, 26,* 183–196.

Souza, P., Arehart, K., Miller, C. W., & Muralimanohar, R. K. (2011). Effects of age on F0- discrimination and intonation perception in simulated electric and electro-acoustic hearing. *Ear and Hearing, 32*(1), 75–83.

Spitzer, S. M., Liss, J. M., & Mattys, S. L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *The Journal of the Acoustical Society of America, 122,* 3678–3687.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2,* 191–196.

Summers, V., & Leek, M. R. (1998). F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *Journal of Speech, Language, and Hearing Research, 41,* 1294–1306.

Takahashi, G. A., & Bacon, S. P. (1992). Modulation detection, modulation masking, and speech understanding in noise in the elderly. *Journal of Speech and Hearing Research, 35,* 1410–1421.

Tillman, T. W., & Carhart, R. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6.* Unpublished manuscript, Northwestern University, Auditory Research Lab, Evanston, IL.

Wang, W. S.-Y. (1967). Phonological features of tone. *International Journal of American Linguistics, 33,* 93–105.

Watson, P. J., & Schlauch, R. S. (2008). The effect of fundamental frequency on the intelligibility of speech with flattened intonation contours. *American Journal of Speech-Language Pathology, 17,* 348–355.

Wilson, R. H., McArdle, R., Betancourt, M. B., Herring, K., Lipton, T., & Chisolm, T. H. (2010). Word-recognition performance in interrupted noise by young listeners with normal hearing and older listeners with hearing loss. *Journal of the American Academy of Audiology, 21,* 90–109.

Wingfield, A., Lombardi, L., & Sokol, S. (1984). Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation. *Journal of Speech and Hearing Research, 27,* 128–134.

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.

Wingfield, A., Wayland, S. C., & Stine, E. A. L. (1992). Adult age differences in the use of prosody for syntactic parsing and recall of spoken sentences. *Journal of Gerontology, 47*(5), P350–P356.

Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., & Rönnberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience, 8,* 88.