

# Effects of fundamental frequency and vocal-tract length cues on sentence segregation by listeners with hearing loss

Carol L. Mackersie,<sup>a)</sup> James Dewey,<sup>b)</sup> and Lesli A. Guthrie

*School of Speech, Language and Hearing Sciences, San Diego State University, 5500 Campanile Drive, San Diego, California 92182-1518*

(Received 15 April 2010; revised 22 May 2011; accepted 1 June 2011)

The purpose was to determine the effect of hearing loss on the ability to separate competing talkers using talker differences in fundamental frequency (F0) and apparent vocal-tract length (VTL). Performance of 13 adults with hearing loss and 6 adults with normal hearing was measured using the Coordinate Response Measure. For listeners with hearing loss, the speech was amplified and filtered according to the NAL-RP hearing aid prescription. Target-to-competition ratios varied from 0 to 9 dB. The target sentence was randomly assigned to the higher or lower values of F0 or VTL on each trial. Performance improved for F0 differences up to 9 and 6 semitones for people with normal hearing and hearing loss, respectively, but only when the target talker had the higher F0. Recognition for the lower F0 target improved when trial-to-trial uncertainty was removed (9-semitone condition). Scores improved with increasing differences in VTL for the normal-hearing group. On average, hearing-impaired listeners did not benefit from VTL cues, but substantial inter-subject variability was observed. The amount of benefit from VTL cues was related to the average hearing loss in the 1–3-kHz region when the target talker had the shorter VTL. © 2011 Acoustical Society of America. [DOI: 10.1121/1.3605548]

PACS number(s): 43.71.Ky, 43.71.Bp, 43.66.Sr, 43.66.Ts [BCM]

Pages: 1006–1019

## I. INTRODUCTION

The reduction in speech understanding in the presence of competing backgrounds depends, in part, on the nature of the background. When the competition consists of spectrally dense noise or unintelligible babble, speech understanding is largely driven by energetic masking caused by the spectral and temporal overlap of the target and competing signals (Assmann and Summerfield, 2004). A second form of masking, known as “informational masking,” may occur as a result of perceptual confusion caused by signal or masker uncertainty or by linguistic interference (Durlach *et al.*, 2003; Schneider *et al.*, 2007; Watson, 2005). When the competing signal consists of a single intelligible talker, the dominant source of interference is informational rather than energetic masking (Brungart, 2001; Brungart *et al.*, 2006).

Acoustic differences between talkers can substantially enhance the recognition of one talker in the presence of a single competing talker, presumably by reducing informational masking (Bregman, 1990; Schneider *et al.*, 2007). It is easier, for example, to recognize a talker when the competing talker has a different gender than when the competing talker has the same gender (Brungart, 2001; Brungart *et al.*, 2001). The two primary differences between talkers of different gender, fundamental frequency (F0) and vocal-tract length (VTL), have been shown to provide robust acoustic cues for perceptual segregation (Darwin *et al.*, 2003; Vestergaard *et al.*, 2009; Vestergaard and Patterson, 2009). The robustness of these

cues for people with cochlear hearing loss is less clear, however. The use of F0 and VTL segregation cues by listeners with hearing loss was the focus of the present study.

### A. Fundamental frequency cues

The F0 and harmonically related components of a talker’s speech convey pitch information and contribute to perceived gender, age and size (Smith and Patterson, 2005). Improvement in intelligibility with increasing talker differences in F0 has been observed for both double-vowel stimuli and sentences (Assmann, 1999; Bird and Darwin, 1998; Brox and Nootboom, 1982; Darwin *et al.*, 2003). For listeners with normal hearing, improvement in sentence recognition occurs for F0 differences up to 10 to 12 semitones (Bird and Darwin, 1998; Darwin *et al.*, 2003; Drullman and Bronkhorst, 2004).

There is evidence that some listeners with hearing loss are unable to take full advantage of F0 differences between target and competing speech. In a series of studies examining the effectiveness of F0 cues in double-vowel perception, Arehart and her colleagues consistently demonstrated that individuals with hearing loss were less able to use F0 cues than were listeners with normal hearing (Arehart, 1998; Arehart *et al.*, 1997, 2005). Differences between normal-hearing and hearing-impaired listeners have also been observed for sentence materials. Summers and Leek (1998) reported that benefit from increasing F0 differences between two sentences was similar for normal-hearing and hearing-impaired listeners for small F0 differences (2 semitones), but some hearing-impaired listeners were unable to take advantage of a larger F0 difference (4 semitones).

The reduction in F0 benefit among hearing-impaired listeners parallels the impaired pitch discrimination observed

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: cmackers@mail.sdsu.edu

<sup>b)</sup>Current address: Speech and Hearing Sciences, Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University 2240 Campus Drive, Evanston, IL 60208.

for other complex stimuli (Arehart, 1994; Bernstein and Oxenham, 2006; Leek and Summers, 2001; Moore and Peters, 1992). Similarly, hearing-impaired listeners have greater difficulty than do normal-hearing listeners in perceptual segregation of competing melodies and sequences of pure tones or harmonic complexes (Grimault *et al.*, 2000; Grose and Hall, 1996; Mackersie *et al.*, 2001; Rose and Moore, 1997, 2000).

The reduction in the ability of listeners with hearing loss to use talker differences in F0 may result from weaker than normal pitch perception caused by poor resolution of harmonics or a reduction in the ability to use periodicity information. In addition, cognitive and other age-related factors may be involved, as older listeners tend to show less benefit from F0-difference cues than do younger listeners (Rossi-Katz and Arehart, 2009; Summers and Leek, 1998). Inadequate audibility of high-frequency harmonics may also weaken auditory grouping based on harmonic structure. Low-frequency harmonics have generally been shown to dominate in the perception of pitch (Dai, 2000; Moore *et al.*, 1985), but listeners appear to use information across a broader range of frequencies to facilitate the perceptual grouping of harmonics in both vowels (Culling and Darwin, 1993; Rossi-Katz and Arehart, 2005) and sentences (Bird and Darwin, 1998). Bird and Darwin, for example, determined that listeners with normal hearing used both high- and low-frequency information in sentences when the talker differences in F0 were large (five semitones or more), but not when F0 differences were small (two semitones and lower). It is, therefore, possible that hearing-impaired listeners' limited access to high-frequency information partially explains the plateau in performance observed for larger F0 differences in the study conducted by Summers and Leek (1998).

In studies examining specific talker segregation cues, individualized amplification and frequency shaping have rarely been used to test listeners with hearing loss. Arehart *et al.* (1998) reported that amplification and low-pass filtering of stimuli in a double-vowel experiment did not increase the benefit from F0 differences over that observed for unfiltered stimuli. It is possible, however, that the single combination of filtering and amplification did not optimize the response for all listeners. In a later study in which the authors used individualized amplification, benefit from F0 differences in vowels was similar for listeners with and without hearing loss (Rossi-Katz and Arehart, 2005). Given the weak connection between double-vowel and sentence recognition (Summers and Leek, 1998); however, it is unknown to what extent these results would apply to sentences.

## B. Vocal-tract length cues

Vocal-tract length affects the average spectral envelope and formant frequencies of a talker's speech. Perception of talker size depends on VTL cues (Ives *et al.*, 2005; Smith and Patterson, 2005). In addition, VTL cues have been shown to influence judgments of talker age and gender (Smith and Patterson, 2005). For listeners with normal hearing, increasing the difference between the average spectral envelope of a target and competing sentence results in an

improvement in speech recognition (Darwin *et al.*, 2003; Darwin and Hukin, 2000a, 2000b; Vestergaard *et al.*, 2009). Darwin and his colleagues reported improvements of more than 20 percentage points as the VTL ratios were increased from 1.0 to 1.38. Also, the combined effects of F0 and VTL cues resulted in greater improvement than either cue alone.

The relative contribution of F0 and VTL cues was examined by Vestergaard *et al.* (2009) by co-varying VTL and glottal pulse rate (associated with F0) over a wide range. They matched the amplitude envelopes of target and competing syllables to prevent listeners from listening in the dips of the envelopes of the competing sounds. A change in VTL of 1.6 times the glottal pulse rate was needed to equate performance for VTL and glottal pulse rate manipulations. A similar trading relationship was found with simulated spatial separations of up to 8° azimuth (Ives *et al.*, 2010). These findings suggest that although both pitch and VTL cues contribute substantially to talker segregation for normal-hearing listeners, pitch is a more prominent cue.

The extent to which listeners with hearing loss can use vocal tract length cues for talker segregation is unknown. If, however, this ability requires access to at least the first two formants, then it is possible that a reduction in high-frequency audibility and/or frequency resolution typically associated with sensorineural hearing loss will affect performance, even when using appropriate amplification.

The primary purpose of the present study was to determine how well listeners with cochlear hearing loss can use F0 and apparent VTL cues to segregate two competing sentences when provided with individualized amplification. The talker cues were manipulated independently.

A secondary purpose was to evaluate the effects of hearing loss on the relative influence of energetic versus informational masking by comparing sentence recognition in the presence of a competing sentence and amplitude-modulated (AM) noise. Using CRM sentences with limited linguistic cues, Brungart (2001a) showed that recognition by normal-hearing listeners was substantially poorer in the presence of a single competing sentence than in the presence of more spectrally dense AM noise. Brungart interpreted the poorer performance with a single competing talker as evidence that performance in this condition was limited primarily by informational masking. However, energetic masking may have a greater influence on performance of listeners with hearing loss due to impaired frequency and temporal resolution. In the present study, the amplitude envelopes of individual CRM competing sentences were extracted and used to modulate noise so that energetic masking could be estimated for AM noise matched to the envelope of the competing sentence. It is important to note, however, that in different frequency bands, representation of the AM noise and speech envelopes within the auditory system would not be the same; auditory representation of the speech envelope would vary across frequency, whereas the AM noise envelope would not.

Listeners with hearing loss were provided with individualized amplification that mimicked the frequency response of a hearing aid. Although the frequency response of a typical hearing aid does not entirely compensate for the loss of

audibility, the use of speech filtered to mimic a hearing-aid response may provide a clearer picture of the accessibility of talker difference cues for typical hearing aid users. Sentences were filtered and amplified to approximate the frequency response prescribed by the NAL-RP hearing aid prescription (Byrne and Cotton, 1988), which is designed to equalize loudness across mid-range frequencies.

## II. GENERAL METHOD

### A. Listeners

Thirteen adults with sensorineural hearing loss and six adults with normal hearing participated in the study. Table I shows the pure-tone audiometric thresholds and monosyllabic word-recognition scores in quiet for the listeners' test ears.

The mean three-frequency average threshold (0.5, 1, and 2 kHz) of listeners with hearing loss was 46 dB hearing level and the mean monosyllabic word recognition score in quiet was 81%. Listeners had normal middle-ear admittance and air-bone gaps of 10 dB or less between 500 and 4000 Hz, indicating that there was no conductive involvement. Acoustic reflex findings for listeners with sensorineural hearing loss were within the 90% confidence intervals for cochlear hearing loss (Silman and Gelfand, 1981), suggesting that there was no retrocochlear involvement. The etiologies of the hearing losses varied: six listeners reported a history of noise exposure, one reported a family history of hearing loss, and two had suspected vascular etiologies. The remaining four had unknown etiologies.

Listeners with normal hearing ranged in age from 25 to 69 yr (mean: 48 yr), whereas listeners with hearing loss

TABLE I. Age in years, monosyllabic word-recognition scores (WS%), pure-tone averages (PTA) in dB HL, and the test-ear pure-tone thresholds from 250 to 8000 Hz (in dB HL) for listeners with hearing loss (first 13) and normal hearing.

ID	Age (yrs)	WS%	PTA (dB HL)	Frequency (Hz)								
				250	500	1000	1500	2000	3000	4000	6000	8000
HL01	61	68	48	25	30	45	60	70	70	70	75	80
HL02	76	77	30	15	15	30	40	45	55	60	75	85
HL03	59	89	40	20	25	35	50	60	65	70	70	75
HL04	67	83	47	25	30	50	50	60	65	65	70	90
HL05	45	73	52	20	25	55	75	75	70	65	85	NR
HL06	59	83	47	25	40	50	45	50	60	55	50	45
HL07	61	87	63	40	60	60	55	70	60	65	70	70
HL08	61	90	38	30	35	40	40	40	45	50	55	60
HL09	68	78	43	35	30	45	50	55	65	70	90	NR
HL10	62	97	22	-5	-5	10	40	60	65	80	80	55
HL11	63	73	45	10	15	40	80	80	85	95	NR	NR
HL12	59	76	60	55	60	60	60	60	65	70	70	70
HL13	55	77	45	15	25	45	55	65	90	95	NR	NR
NH01	68	100	8	0	5	10	10	5	10	5	25	25
NH02	50	100	6	0	0	10	10	5	10	5	0	0
NH03	53	100	14	10	15	20	15	5	15	25	25	25
NH04	52	100	4	0	0	0	5	10	10	5	20	15
NH05	25	100	6	5	10	5	10	0	10	15	10	10
NH06	59	100	15	20	20	15	15	10	20	25	20	20

ranged in age from 45 to 76 yr of age (mean: 61 yr). All listeners were native speakers of American English and were screened for cognitive disabilities using the Mini-Mental Status Examination (Folstein *et al.*, 1975). All listeners scored 29 or 30 out of 30 possible points. The test ear was randomly selected for listeners with bilaterally symmetrical hearing. For one listener with asymmetrical hearing, the better ear was selected; this ear most closely corresponded to an average hearing loss of 50 dB hearing level.

### B. Materials and processing

#### 1. Amplitude-modulated (AM) noise

Amplitude-modulated noise stimuli were created by extracting the amplitude envelopes from each competing sentence and applying them to samples of random noise. The upper frequency limit of the envelopes was 60 Hz. The random noise was filtered to match the average spectrum of five sentences with no F0 or VTL shift. Two different infinite impulse response low-pass filters were used to filter the noise samples. One filter had a cut-off of 625 Hz and a slope of 11 dB/octave (dB/cycle). The second filter had a cut-off of 1250 Hz with a slope of 14 dB/octave (dB per cycle). The noise filtered with the 625 Hz cut-off was combined with the noise filtered with the 1250 Hz cut-off. These operations resulted in a set of amplitude-modulated noise samples, each with amplitude envelopes that matched the single competing sentence from which it was created. The rms level of each noise sample was matched to that of the competing sentences.

#### 2. Speech materials

Speech materials were drawn from the Coordinate Response Measure (CRM) speech corpus developed for multi-talker research (Bolia *et al.*, 2000). Sentences in the corpus are constructed from key words consisting of call signs ("Baron," "Charlie," "Eagle," etc.), colors ("red," "white," "green," and "blue"), and numbers (1–8). These key words are combined with simple cues: "ready," "go to," and "now," to form sentences (e.g., "Ready" + "Charlie" + "go to" + "red" + "one" + "now"). The listener's task is to repeat the key words (color and number) associated with the call sign "Baron" while ignoring the competing sentence. Six of the eight call signs were used ("Baron," "Charlie," "Ringo," "Hopper," "Tiger," "Eagle").

A single male talker (Talker 0) from the Coordinate Response Measure (CRM) speech corpus was used for both the target and competing sentences. Talker differences were created by manipulating the characteristics of the talker's voice to produce changes in F0 and apparent vocal tract length, as described below.

#### 3. Speech Processing

Speech manipulations were made using the pitch-synchronous overlap-add (PSOLA) algorithm (Moulines and Charpentier, 1990) as implemented by the Praat software package (Boersma and Weenink, 2006). Fundamental frequency and VTL were manipulated independently. The F0s



were shifted by  $-3$ ,  $0$ ,  $+3$ , and  $+6$  semitones. Formant frequencies and speaking rate are unaffected by this processing.

The spectral envelope was scaled to produce changes in the apparent VTL. For convenience, the spectral envelope manipulations will be referred to as changes in VTL throughout the paper. Spectral envelopes were manipulated in the manner described by Darwin *et al.* (2003), resulting in proportional changes of 0.84, 0.92, 0.96, 1.00, 1.04, 1.08, and 1.16. Values below 1.0 correspond to shorter VTLs than the original and values above 1.0 correspond to longer VTLs. Even the smallest VTL shift would be expected to produce perceptually salient changes in perceived talker size based on the findings of Ives *et al.* (2005) showing size discrimination performance of 75% and higher for similar shifts in syllabic stimuli. The desired proportional changes (0.84, 0.92, etc.) were used as scaling factors (sf) to shift F0 and duration ( $F0 \times sf$ ;  $\text{duration} \times 1/sf$ ). The stimuli were then re-sampled (original sampling rate of 40 kHz  $\times$  sf) and stored for playback at the original sampling rate. These operations shifted the spectral envelope, but retained the original duration and F0.

The sentences were combined to produce sentence pairs with talker differences in F0 or VTL. The F0 differences and VTL ratios used are shown in Table II. The processing type (F0 shift or proportional change in the spectral envelope) for the individual sentences comprising the sentences pairs is shown in parentheses. Talker differences in F0 ranged from 0 to 9 semitones. The ratio of spectral envelope shifts for the two sentences comprising each pair (VTL ratios) ranged from 1.00 (no shift) to 1.38.

### C. Procedures

All stimuli were routed from a computer to a Tucker-Davis Technologies RX8 24-bit multi I/O processor and presented monaurally through an Etymotic ER4 insert earphone. Listeners were tested in a sound-attenuated booth.

Presentation levels and stimulus filtering parameters were chosen to achieve the following goals: (1) comfortable loudness for all listeners and (2) frequency shaping for listeners with hearing loss that approximates a typical hearing aid. Unfiltered stimuli were presented to normal-hearing listeners at an average level of 71 dB sound pressure level, which corresponded to a comfortable loudness level for all listeners (see description of loudness testing below).

Presentation levels for listeners with hearing loss were based on a three-stage process (described below) involving

TABLE II. Fundamental frequency and vocal-tract length (VTL) manipulations used to produce talker differences. The talker difference (in semitones or VTL ratio) is shown at the left of each column. The shifts for the individual sentences comprising the pairs are shown in parentheses.

F0 difference (shift in semitones)	VTL Ratio (Proportional shift)
0 (0, 0)	1.0 (1.00, 1.00)
3 (-3, 0)	1.08 (0.96, 1.04)
6 (-3, +3)	1.16 (0.92, 1.08)
9 (-3, +6)	1.38 (0.84, 1.16)

(1) amplification and frequency shaping (2) verification of output level needed to adequately match the prescribed NAL-RP targets and (3) verification/adjustment of overall output to the listener's most comfortable loudness level.

### 1. Frequency shaping/amplification and verification

For listeners with hearing loss, stimuli were individually amplified and filtered to approximate the frequency shaping prescribed by the National Acoustics Laboratory revised hearing aid formula (NAL-RP) (Byrne and Cotton, 1988). Individual filters were created for each listener. Amplification and filtering of the speech stimuli were based on targets for a 65 dB speech-shaped noise signal.

Individual real-ear-to-coupler differences (RECD) were measured with the Fonix 7000 using the same insert earphone used for subsequent speech recognition testing. The RECD was used to convert real-ear sound pressure level targets to 2 cc coupler targets.

Frequency responses were verified by comparing the spectrum of the amplified/filtered speech-shaped noise to the 2 cc target value. The amplified/filtered noise was played from the test system (Tucker-Davis Technologies RX8), delivered to the Etymotic ER4 insert earphone attached to a 2cc coupler, and measured using the Frye Fonix 7000 analysis system. The spectrum of the coupler response was compared to the target spectrum. Hearing aid microphone effects were excluded from the calculations. Filters were redesigned as needed until the measured values were as close to the target values as possible. Figure 1 shows the mean target and measured 2 cc coupler outputs (dB sound pressure level in 1/3 octave bands) for the 13 hearing-impaired listeners.

### 2. Loudness ratings

Categorical loudness ratings were obtained to verify that the amplifier and filter settings produced an output that corresponded to a comfortable loudness level. Starting with a sentence at a level 4 or 6 dB lower than the level corresponding

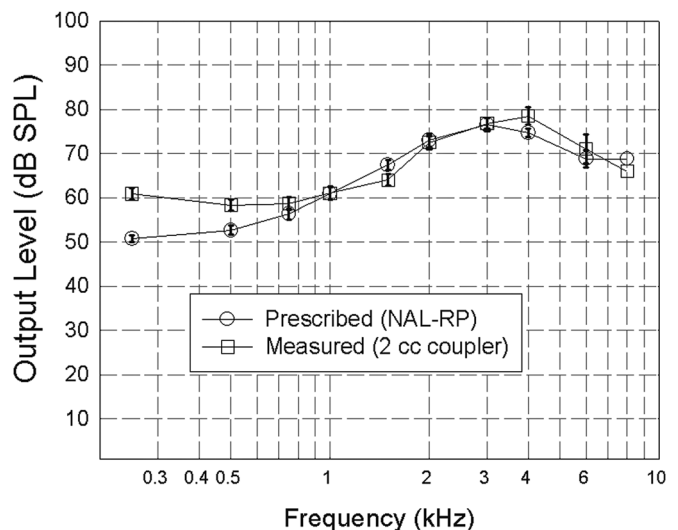


FIG. 1. Mean target and measured 2cc coupler outputs (in dB sound pressure level) for the listeners with hearing loss, measured in 1/3 octave bands.

to the prescribed settings, listeners were asked to rate the loudness on a scale of 2–8: [2—“Very soft”; 3—“Soft”; 4—“Comfortable, but slightly soft”; 5—“Comfortable”; 6—“Comfortable, but loud”; 7—“Loud, but OK”; or 8—“Uncomfortably loud.” The presentation level was increased in 3 dB steps until the listener reported a loudness rating above 5. The procedure was repeated until two ratings of “comfortable” were obtained at the same level. Typically, only one repetition was needed. The final presentation level for the speech stimuli was the level at which listeners consistently reported a loudness rating of “5.” On average, the prescribed and preferred levels were within 1.2 dB. Normal-hearing listeners also completed loudness ratings to ensure that the 71 dB sound pressure level stimuli fell within the “comfortable” range. This process was necessary to minimize possible confounding effects of inappropriate amplification.

## D. Statistical analyses

Recognition scores (percentage correct) were converted to rationalized arcsine units for statistical analyses in order to stabilize the error variance (Studebaker, 1985). For all competing-talker conditions, repeated-measures analyses-of-variance were used to test for differences in means for the factors of interest. Partial eta-squared ( $\eta^2_p$ ) was used as a measure of effect size. Greenhouse-Geisser corrections were applied whenever violations of the sphericity assumption occurred. Newman-Keuls *post hoc* tests were used for pairwise comparisons of significant interactions and/or main effects. A probability of .05 was used as a criterion for statistical significance.

## III. EXPERIMENT 1: TALKER SEGREGATION WITH RANDOMIZATION OF TARGET-TALKER CHARACTERISTICS

### A. Method

Recognition tests were administered across three test sessions of approximately 60 min each. Performance in quiet was measured during the first session. Performance was measured in the presence of both AM noise and competing sentences during sessions two and three, as described below.

#### 1. Performance in quiet

To ensure that the F0 and VTL manipulations did not affect the intelligibility of speech in quiet, recognition of sentences was evaluated under all F0 and VTL processing conditions without competition. Listeners were given five practice blocks, each consisting of 10 sentences (20 key words) to orient them to the test procedure. Practice conditions consisted of the sentences with no shift in F0 or VTL and the sentences with the largest and smallest changes in F0 and VTL.

Practice segments were followed by separate test blocks, each containing ten sentences. The conditions with F0 and VTL shifts were presented in random order for each listener. The F0 and VTL reference conditions (no shift) were tested at the beginning or end of the session, counterbalanced across listeners. Performance was measured as the percentage of correctly repeated items (colors and numbers).

#### 2. Performance in the presence of a competing talker or noise

*a. Practice.* To familiarize listeners with the competing sentence task, initial training was provided for sentences with and without F0 and VTL changes. During the initial training, the target-to-competition ratio (TCR) was decreased from +20 to 0 dB in 5-dB steps across a series of presentations for each condition. Re-instruction and additional practice were given to listeners until they became familiar with the task and performance stabilized. Following this initial training, two practice blocks were administered at a fixed TCR of 0 dB.

*b. Test conditions.* The target sentence (call sign: “Baron”) was combined with a single competing sentence or with AM noise. Only the sentences with no shift in F0 or VTL were combined with the AM noise; sentences for all F0 or VTL shifted conditions were combined with the competing speech. The F0 and VTL conditions were completed in separate test sessions, with the test order counterbalanced within the two groups of listeners. The AM noise conditions were randomly assigned to either the beginning or the end of each session.

Sentences were presented at four TCRs (0, 3, 6, and 9 dB). Each condition (VTL, F0, or AM noise) was completed in a single 40-sentence block (10 sentences for each TCR). Easier TCRs (9 or 6 dB) were always presented first and were alternated with the more difficult TCRs (0 or 3 dB). The TCR sequence was randomly assigned to each listener. All TCRs for a given processing condition were completed before beginning a new processing condition.

During conditions in which there were talker differences, the target and competing talkers were randomly assigned to the higher or lower F0 or VTL values within each block of 10 sentences. The high and low values were equally distributed across the blocks of sentences presented at each TCR.

As in the training session, listeners were asked to repeat the color and number spoken by the talker with the call sign “Baron” following each presentation. The percentage of correctly repeated items (colors and numbers) was recorded.

## B. RESULTS

### 1. Single-sentence performance

For all listeners with normal hearing and 10 of 13 listeners with hearing loss, there were no errors when listening to sentences without competition. For listeners with hearing loss the mean scores were 100, 99.2, 99.6, and 99.6% for F0 shifts of  $-3$ , 0,  $+3$ , and  $+6$  semitones, respectively. Mean scores for the VTL conditions were 98.5, 98.9, 99.2, 98.6, 99.2, and 99.6% for VTL values of 0.84, 0.96, 1.00, 1.04, 1.08, and 1.16, respectively. Skewed distributions precluded the use of parametric statistics. Instead, separate Friedman analysis of variance tests were completed for the F0 and VTL data of listeners with hearing loss. There was no significant effect of processing for either the F0 ( $X^2(3) = 1.40$ ,  $p = 0.70$ ) or VTL conditions ( $X^2(6) = 3.17$ ,  $p = 0.78$ ). Based

TABLE III. Mean recognition scores (% correct) for sentences without a F0 or VTL shift for TCRs of 0, 3, 6, and 9 dB for two masker types: amplitude-modulated (AM) noise and single sentences. Standard deviations are shown in parentheses.

TCR	Normal Hearing		Hearing Loss	
	AM noise	Single sentence	AM noise	Single sentence
0	99.6 (1.1)	55.8 (13.4)	95.0 (8.6)	49.4 (7.9)
3	100 (0)	62.2 (13.9)	96.9 (5.5)	71.1 (12.7)
6	100 (0)	87.5 (14.6)	98.5 (2.6)	83.2 (10.0)
9	100 (0)	98.7 (2.2)	98.8 (2.9)	93.2 (7.5)

on these analyses, there is no evidence that the F0 or VTL processing had a substantial effect on speech intelligibility in quiet.

## 2. Effect of amplitude-modulated noise and a competing talker

Table III shows mean recognition scores for sentences tested in the presence of AM noise and single competing sentences. Scores represent the means for sentences with no shift in VTL or F0. Mean scores in AM noise were close to 100% for all TCRs. All but two listeners with hearing loss scored 95% or higher under all AM noise conditions. Scores in the presence of a single competing sentence were substantially lower than scores in AM noise for both listener groups. This finding suggests that, even for listeners with hearing loss, informational masking was the dominant factor underlying performance in the competing talker conditions for the TCRs used in this study.

## 3. Effects of talker differences in F0

*a. Performance for high- and low-F0 target talkers.* Mean scores for high- and low-F0 target talkers are shown in Figs. 2 and 3, respectively. When the target talker had the higher F0 value (Fig. 2), recognition improved with increasing F0 difference for both groups, mainly for the lower TCRs. At lower TCRs, listeners with normal hearing benefited more from a nine semitone F0 difference than did listeners with hearing loss. For the higher TCRs, there was less room for improvement as performance approached ceiling level.

When the target talker had the lower F0 (Fig. 3), there was no evidence of benefit from increasing F0 difference for either group. At the lowest TCR, mean performance of listeners with hearing loss decreased with increasing difference in F0.

Performance for the high and low target talkers was analyzed separately. A repeated-measures analysis of variance was completed using hearing status as a between-subjects factor, and F0 difference (0, 3, 6, 9 semitones) and TCR (0, +3 dB) as within-subjects factors. Target-to-competition ratios of +6 and +9 dB were not included because at the higher TCRs, more than half of the listeners approached ceiling-level performance (> 90%) for the reference condition.

*b. Analysis of variance: High-F0 target.* A summary of the analysis of variance results is shown in Table IV. There were significant interactions between F0 difference and hearing status and between TCR and hearing status. *Post hoc* testing confirmed a significant improvement in recognition from 6 to 9 semitones for listeners with normal hearing

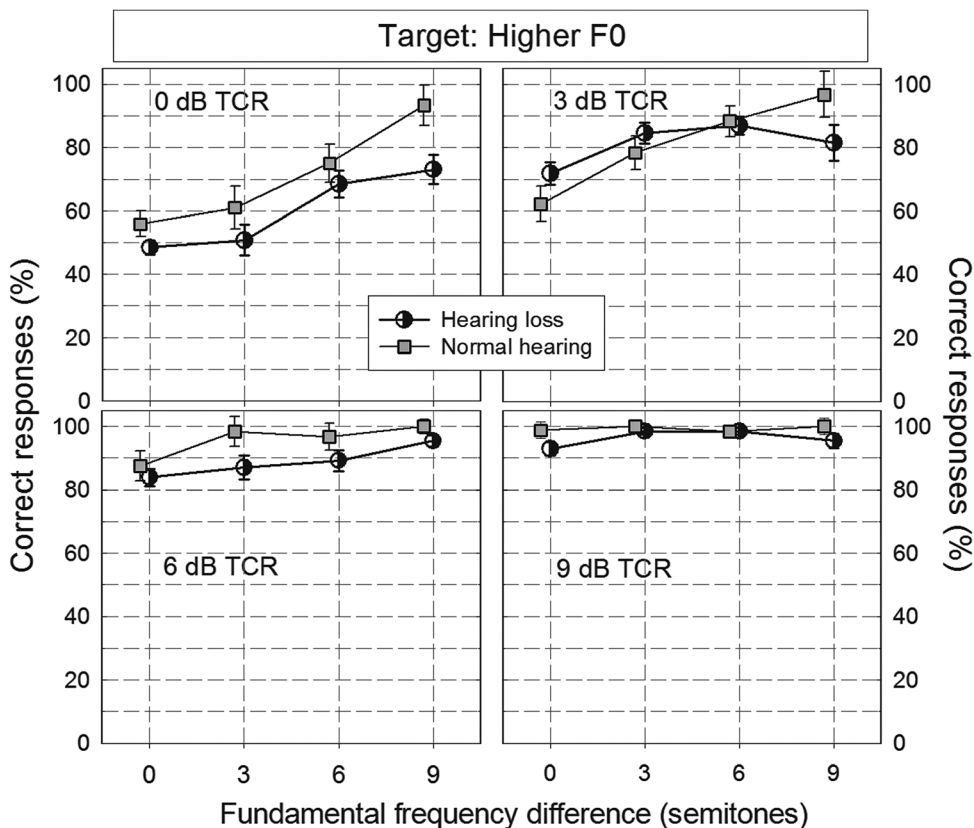


FIG. 2. Mean scores for trials in which the target was assigned to the higher F0 value. Data for different TCRs are shown in separate panels. In this and subsequent figures, error bars indicate  $\pm 1$  standard error.

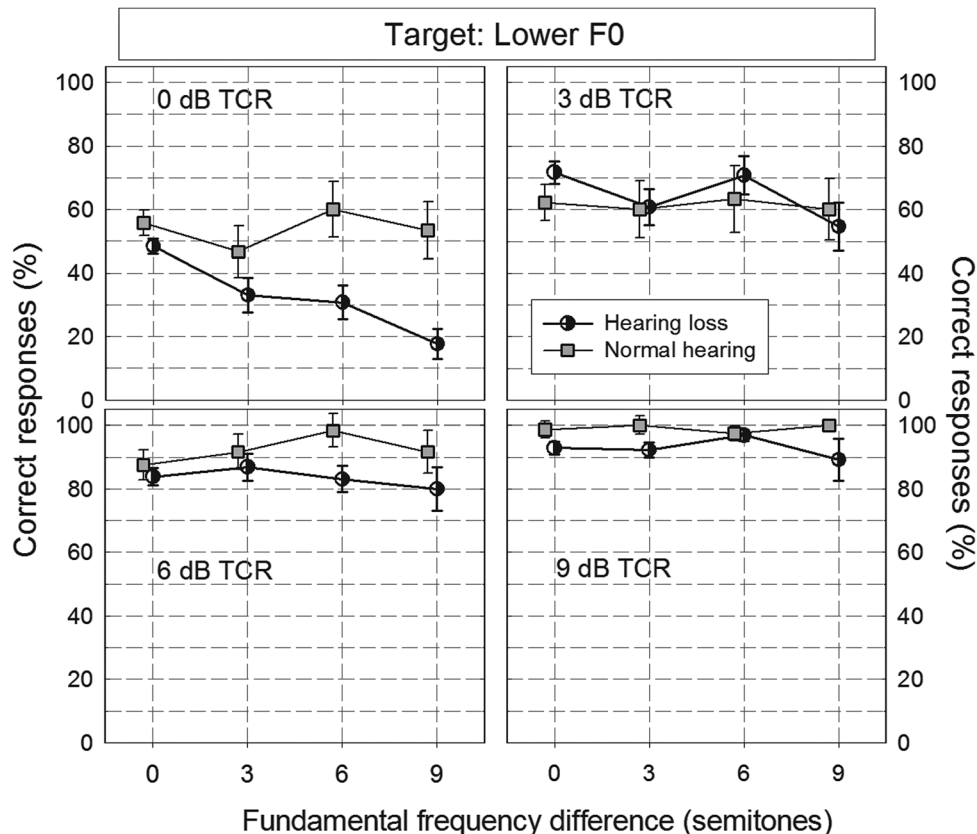


FIG. 3. Mean scores for trials in which the target was assigned to the lower F0 value.

( $p = 0.004$ ), but not for listeners with hearing loss ( $p = 0.92$ ). For both groups, scores for F0 differences of 6 and 9 semitones were significantly higher than scores for an F0 difference of 0 semitones. The mean score for normal-hearing listeners was significantly higher than the score for listeners with hearing loss ( $p = 0.006$ ) for a difference of nine semitones. The mean score for listeners with hearing loss was poorer than the score for normal-hearing listeners at the lower TCR, but was similar at the higher TCR.

*c. Analysis of variance: Low F0 target.* There was no significant effect of F0 difference or interaction involving F0

TABLE IV. Analyses of variance results for high- and low-F0 targets. Significant  $p$  values are shown in bold.

Effect: High F0 target	df	F	$p$	$\eta^2_p$
Hearing	(1,17)	4.08	0.059	0.193
F0 difference	(3,51)	17.21	<b>&lt;0.001</b>	0.503
TCR	(1,17)	43.82	<b>&lt;0.001</b>	0.720
F0 diff $\times$ Hearing	(3,51)	3.05	<b>0.037</b>	0.152
F0 diff $\times$ TCR	(3,51)	2.86	0.064	0.144
TCR $\times$ Hearing	(1,17)	4.86	<b>0.042</b>	0.222
F0 diff $\times$ TCR $\times$ Hearing	(3,51)	0.41	0.748	0.023
Effect: Low F0 target	df	F	$p$	$\eta^2_p$
Hearing	(1,17)	2.52	0.131	0.129
F0 difference	(3,54)	1.63	0.194	0.087
TCR	(1,17)	34.98	<b>&lt;0.001</b>	0.673
F0 diff $\times$ Hearing	(3,54)	2.25	0.093	0.117
F0 diff $\times$ TCR	(3,54)	0.85	0.471	0.048
TCR $\times$ Hearing	(1,17)	10.35	<b>0.005</b>	0.378
F0 diff $\times$ TCR $\times$ Hearing	(3,54)	1.20	0.319	0.066

difference, confirming that F0 difference did not enhance recognition for the low F0 target. As in the previous analysis, the significant interaction between TCR and hearing status reflected greater differences between groups for the lower TCR. There were no other main effects or interactions.

#### 4. Effects of talker differences in apparent vocal tract length

Mean scores for the high- and low-VTL targets were 77.1% and 78.0%, respectively. A repeated-measures analysis of variance did not reveal a significant difference between high and low target values ( $F(1,17) = 0.58$ ,  $p = 0.45$ ) nor any significant interactions between VTL target and any other factor. Therefore, scores for the high and low VTLs were averaged for subsequent analysis.

Mean scores are shown in Fig. 4 as a function of VTL ratio. On average, recognition by normal-hearing listeners improved with increasing VTL ratio. This effect was absent for the group with hearing loss.

A repeated-measures analysis of variance with hearing status as a between-subjects factor and VTL ratio and TCR (0, +3 dB) as within-subjects factors indicated a significant interaction between VTL ratio and hearing status ( $F(3,51) = 12.29$ ,  $p < 0.0001$ ,  $\eta^2_p = 0.42$ ) and between TCR and hearing status ( $F(3,51) = 4.62$ ,  $p = 0.046$ ,  $\eta^2_p = 0.214$ ). *Post hoc* tests for the normal-hearing group confirmed a significant improvement in scores (re no VTL shift) for VTL ratios of 1.16 ( $p = 0.02$ ) and 1.38 ( $p = 0.0002$ ), but no significant improvement for a VTL ratio of 1.08 ( $p = 0.65$ ). In contrast, there was no significant improvement in scores for the group with hearing loss for any VTL ratio. These data provide no



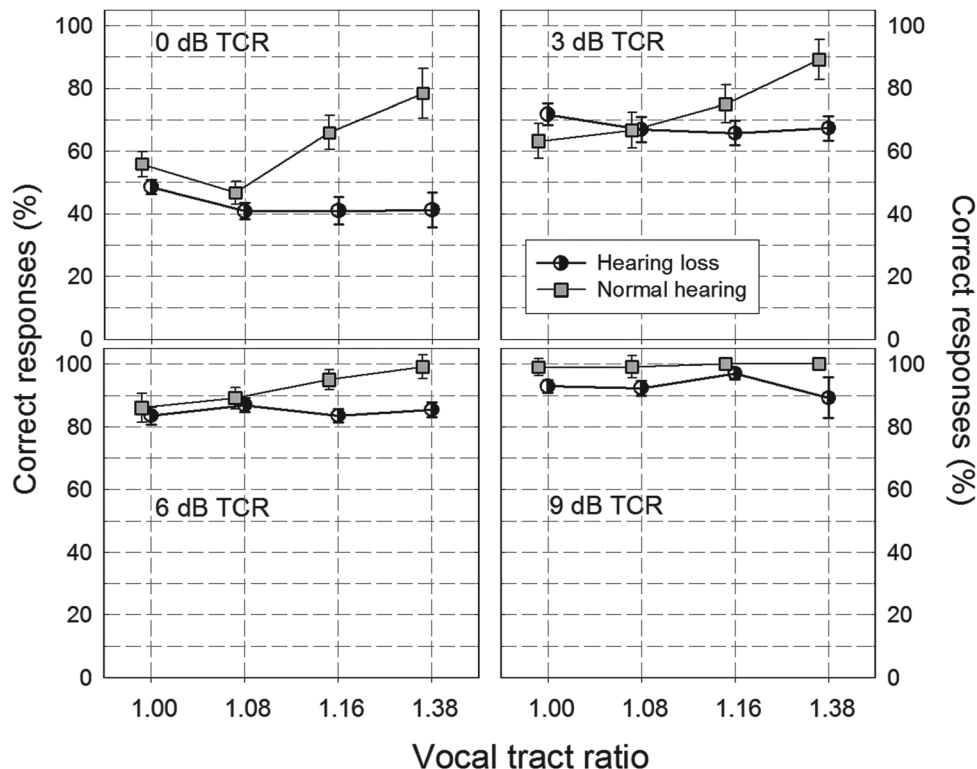


FIG. 4. Mean scores under conditions in which talker differences in VTL varied from 1.0 to 1.38. Data were averaged across the high and low target-talker values.

evidence that listeners with hearing loss can use differences of vocal-tract length to aid in the separation of two competing talkers, even when using appropriate amplification.

### 5. Summary

Both groups of listeners benefited from talker difference in F0 for the higher-F0 target, but not for the lower-F0 target. Recognition scores were significantly better for the higher-F0 target than for the lower-F0 target. The striking asymmetry between results for the higher- and lower-F0 targets suggests that listeners had difficulty focusing on the lower pitch in the presence of the higher-F0 competition. Scores for the higher- and lower-VTL targets were similar. Listeners with normal hearing benefited from VTL cues, but listeners with hearing loss did not. Recall that in both the F0 and VTL segments of the experiment, the higher and lower targets were randomly selected on each trial. It is possible that the demands of monitoring and switching attention limited performance under some conditions. This possibility was evaluated in Experiment 2.

## IV. EXPERIMENT 2: ELIMINATION OF TRIAL-TO-TRIAL UNCERTAINTY

Experiment 2 was conducted to examine the possible influence of trial-to-trial target uncertainty on the results of the Experiment 1. The goals were (1) to determine if target uncertainty affected the benefit from talker differences observed in Experiment 1 and (2) to quantify the contribution of target uncertainty to the high-low asymmetry effects observed for the F0 conditions in Experiment 1. Testing was limited to the most extreme F0 and VTL differences, because any changes in benefit would most likely occur for larger talker differences. Only the more difficult TCRs were

used because performance for the easier TCRs approached ceiling level for Experiment 1.

### A. Method

#### 1. Listeners

The 20 listeners were the same as for Experiment 1.

#### 2. Stimuli and conditions

Listeners were tested with CRM sentences with no average F0 or VTL differences, with a mean F0 difference of 9 semitones, and with a VTL ratio of 1.38. In contrast to Experiment 1, the assignment of the target “Baron” to the higher or lower F0 or VTL was fixed for each block of 10 sentences. Listeners were tested at TCRs of 0 and + 3 dB.

#### 3. Procedures

The training procedure was the same as for Experiment 1. Following training, half the listeners completed the F0 conditions first (both TCRs) and the remainder completed the VTL conditions first. The conditions with no average F0/VTL difference were tested in the middle of the session. The order of the target talker (fixed high, fixed low) was counterbalanced among the listeners. As in Experiment 1, each processing condition was tested in a single block (20 sentences: 10 sentences at each TCR). The order of the TCR conditions was randomly chosen. Listener instructions and scoring were the same as for Experiment 1.

### B. Results

#### 1. Fixed versus random F0 targets

Mean scores for the conditions with no average F0/VTL difference were 57.6% and 59.5% in Experiments 1 and 2,



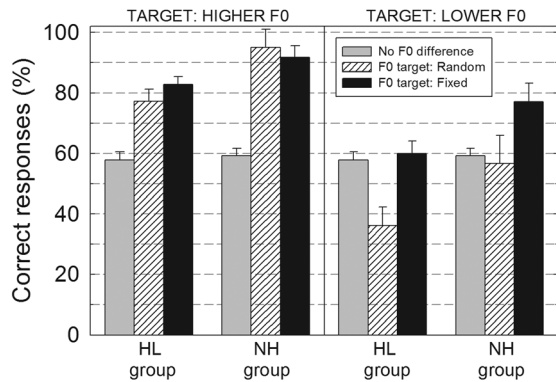


FIG. 5. Mean scores for talker differences of 0 (Experiment 2) and 9 semitones. Scores for a 9 semitone difference are shown for the random- and fixed-target F0 conditions. Scores for the (left) high and (right) low target F0s are shown. Data represent scores averaged across the 0 and +3 dB TCRs.

respectively; this difference was not significant ( $F(1,17) = 0.53, p = 0.48$ ). Therefore, scores for conditions with no average F0/VTL difference were collapsed across Experiments 1 and 2 for analysis.

Scores for F0 differences of 0 and 9 semitones are shown in Fig. 5 for the random-F0 target (Experiment 1) and the fixed-F0 target. Data points represent scores averaged across the 0 and +3 dB TCRs. Mean scores for the random and fixed high-F0 targets were 77% and 83% for the listeners with hearing loss and 95% and 92% for the listeners with normal hearing, indicating that there was little improvement in mean scores for either group when target uncertainty was eliminated. For the low-F0 targets (right), however, there was substantial improvement when target uncertainty was eliminated. Mean scores for the random and fixed low-F0 targets were 36% and 60% for the listeners with hearing loss and 57% and 77% for the listeners with normal hearing. Relative to no F0 difference, mean scores for the fixed low-F0 target were higher for listeners with normal hearing, but not for listeners with hearing loss.

Separate analyses of variance were completed for the high- and low-target conditions using hearing status as a between-subjects factor and processing/presentation (no F0 difference, fixed, random) and TCR (0, 3 dB) as within-subjects factors.

**a. Analysis of variance: High-F0 target.** There was a significant interaction between processing/presentation and hearing status ( $F(2,34) = 4.77, p = 0.02, \eta_p^2 = 0.22$ ), between TCR and hearing status ( $F(1,17) = 9.18, p = 0.008, \eta_p^2 = 0.35$ ), and between processing/presentation and TCR ( $F(2,34) = 0.84, p = 0.43, \eta_p^2 = 0.047$ ). There was no significant interaction between processing/presentation, TCR, and hearing status ( $F(2,34) = 0.12, p = 0.89, \eta_p^2 = 0.007$ ). *Post hoc* tests confirmed that there was no significant difference between scores for the fixed and random conditions for either group. Mean scores for hearing-impaired listeners were significantly lower than scores for normal-hearing listeners at the lower TCR, but not at the higher TCR.

**b. Analysis of variance: Low-F0 target.** There was a significant main effect of processing/presentation ( $F(2,34)$

$= 6.91, p = 0.003, \eta_p^2 = 0.28$ ) and weak evidence of an interaction between processing/presentation and hearing status ( $F(2,34) = 2.64, p = 0.085, \eta_p^2 = 0.13$ ). There was also a significant interaction between TCR and hearing status ( $F(1,17) = 7.55, p = 0.013, \eta_p^2 = 0.31$ ). *Post hoc* testing indicated that scores for the fixed conditions were significantly higher than scores for the random conditions for listeners with hearing loss ( $p < 0.013$ ) and marginally higher for listeners with normal hearing ( $p = 0.053$ ). There was no significant benefit from F0 cues for either group, even for the fixed condition.

## 2. Fixed versus random VTL-length targets

Scores for VTL ratios of 1.0 and 1.38 are shown in Fig. 6 for the random and fixed VTL targets. As in Experiment 1, data were averaged across the high- and low-target conditions. When target VTL was fixed, mean scores for a VTL ratio of 1.38 improved by 13 percentage points (54% random, 67% fixed) for listeners with hearing loss and by 10 percentage points for listeners with normal hearing (83% random, 93% fixed).

A repeated-measures analysis of variance revealed a significant interaction between processing/presentation and hearing status ( $F(2,34) = 11.73, p = 0.0001, \eta_p^2 = 0.41$ ) and between TCR and hearing status ( $F(1,17) = 4.64, p = 0.046, \eta_p^2 = 0.21$ ). *Post hoc* tests indicated that scores for the fixed conditions were significantly higher than scores for the random conditions for both groups. Listeners with hearing loss did not show significant benefit from VTL cues, even when the uncertainty was removed ( $p = 0.20$ ), whereas listeners with normal hearing showed benefit from VTL cues for both the fixed and random targets ( $p < 0.0002$ ).

## 3. Summary

Trial-to-trial uncertainty in the target F0 clearly affected performance, but only for the conditions in which the target was assigned to the lower F0 value. Although performance improved when uncertainty was removed, substantial asymmetries between scores for the high- and low-F0 targets remained for both groups. Target uncertainty did not appear to be a major factor limiting the use of VTL cues by listeners with hearing loss. On average, listeners with hearing loss did

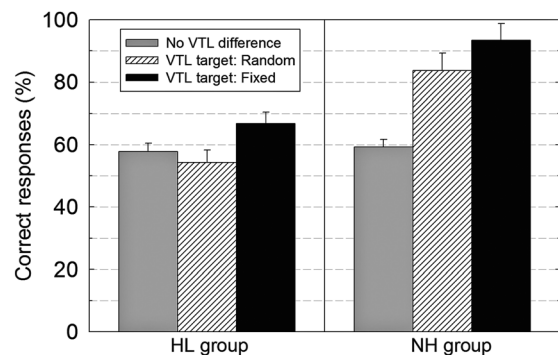


FIG. 6. Mean recognition scores for VTL ratios of 1.0 and 1.38. Scores for the VTL of 1.38 are shown for the random and fixed-target VTL conditions. Data represent scores averaged across the 0 and +3 dB TCRs.

not benefit from talker differences in VTL, even after removal of uncertainty about the target talker's voice.

## V. INDIVIDUAL DIFFERENCES AMONG LISTENERS WITH HEARING LOSS

### A. Performance for high versus low target talker values

The improvements in recognition (re unshifted sentences) for listeners with hearing loss are shown in Fig. 7 for the high- versus low-target F0 and VTL. The fixed-presentation conditions from Experiment 2 are shown. For the F0 conditions (left panel), most listeners showed better performance when the target talker was assigned to the high F0 value. No listener showed better performance for the low F0 value.

The improvement in recognition for the VTL conditions was, on average, similar for the high and low target values. It is apparent from Fig. 7, however, that some listeners performed better with either the longer VTL target or the shorter VTL target. The range in performance was considerably greater for the VTL conditions than for the F0 conditions. There are several instances in which the improvement values were negative (listeners H01, H05, H07), indicating that scores were poorer for the shorter VTL (or lower F0) than for the reference condition. These negative values cannot be explained by differences in audibility because speech intelligibility indices (ANSI, 1997) for the longer and shorter VTL targets were within 0.01 of one another (H01: SII (long, short VTL) = 0.73, 0.72; H05: SII = 0.67, 0.67; H07: SII = 0.68, 0.67). Also, all stimuli were highly intelligible in isolation and energetic masking was minimal for the TCRs used in this study. Therefore, it seems likely that these poor scores occurred as a result of confusion between the lower VTL target and higher VTL competing talkers.

### B. Factors affecting individual differences

Multiple regression analyses were used to evaluate possible sources of individual variability in performance for listeners with hearing loss. The magnitudes of improvements for the fixed conditions (largest F0 difference and VTL ratio) were used as dependent variables. The improvement was averaged across the 0 and 3 dB TCRs. The predictor variables

were listener age, average low-frequency thresholds (0.25, 0.50 kHz) and average mid/high-frequency thresholds (1–3 kHz).

The results of the regression analyses are shown in Table V. The mean 1–3 kHz threshold significantly predicted the improvement for the VTL ratio of 1.38 when the target had the shorter VTL. The adjusted  $R$ -squared value of 0.54 indicates that the absolute threshold explained more than 50% of the variance. As shown in Fig. 8, better hearing in the 1–3 kHz range was associated with greater improvement from VTL processing when the target talker was assigned to the shorter VTL (i.e., more female-like). No listener with average mid/high frequency hearing loss greater than 55 dB hearing level showed improvement with VTL difference. There were, however, several listeners with 1–3 kHz hearing loss less than 60 dB hearing level who showed substantial benefit from VTL differences. As shown in Table V, the independent variables did not significantly predict improvement for the other conditions.

## VI. GENERAL DISCUSSION

### A. Effects of speech competition versus AM noise

Mean recognition scores in AM noise were 95% or higher for both listener groups. As noted earlier, although the broadband envelopes of the competing speech and AM noise were the same, the envelopes of the speech and AM noise, as represented by the auditory system, would differ. Therefore, the opportunities for listening in the dips of the envelope would be different for the two competing signals. However, given the greater spectral density of the AM noise, it is expected that the AM noise would approach the maximum energetic masking possible for the competing sentences. Consistent with previous studies using the CRM materials (Brungart, 2001; Brungart *et al.*, 2001), scores in the presence of the competing sentences were substantially lower than scores in AM noise, particularly for the two lowest TCRs. However, the opposite pattern has been found for sentences with more linguistic context (Festen and Plomp, 1990; Peters *et al.*, 1998), presumably because linguistic cues contributed to the release from information masking produced by the competing speech. These patterns suggest that the relationship between performance with AM noise and a single competing sentence is at least partially

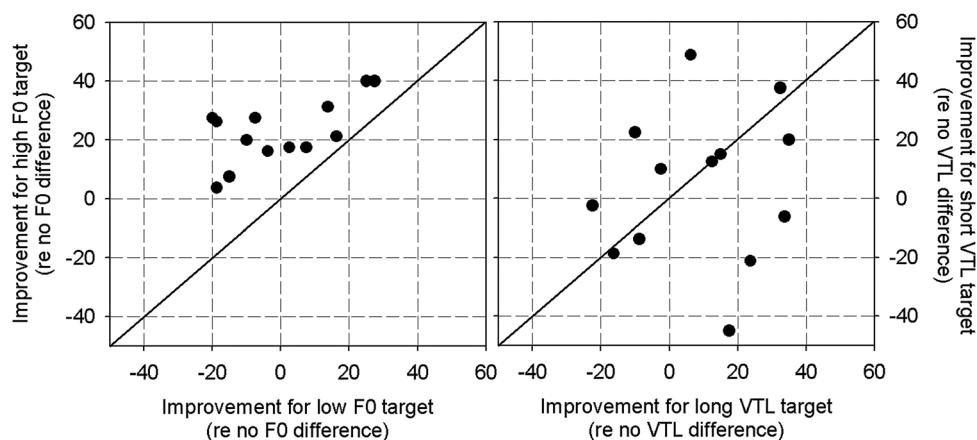


FIG. 7. Improvement in scores (re no average difference in F0/VTL) for listeners with hearing loss (right) for a talker difference of 9 semi-tones (left) and a VTL ratio of 1.38 for blocks in which the target was assigned to the higher (y axis) and lower (x axis) F0 target values and shorter (y axis) and longer (x axis) VTL. Data represent scores averaged across the 0 and +3 dB TCRs.

TABLE V. Results of multiple regression analyses using the improvement in scores (re VTL ratio = 1.0) as the dependent variable and age, average low-frequency hearing loss (0.25–0.5 kHz) and average mid/high-frequency loss (1–3 kHz) as independent variables. The abbreviations “*B*” and “ $\beta$ ” denote unstandardized and standardized beta coefficients, respectively. The adjusted  $R^2$  is shown in parentheses. Separate analyses were completed for talker differences in VTL and F0, and low and high target talkers. The one significant predictor variables is indicated by an asterisk.

	<i>B</i>	SE <i>B</i>	$\beta$	p-level
VTL1.38 Low target (short VT)				
Age	1.28	0.80	0.36	0.31
LF 0.25-0.5	0.31	0.32	0.19	0.11
HF1-3	-1.70	0.66	-0.60	<0.03*
$R^2$ (Adjusted $R^2$ )	0.65 (0.54)			
VTL1.38 High target (long VT)				
Age	0.97	-0.52	-0.18	0.62
LF 0.25-0.5	0.39	0.45	0.15	0.66
HF1-3	0.80	-1.08	-0.39	0.31
$R^2$ (Adjusted $R^2$ )	0.12(-0.18)			
F0 - 9 semitones Low target				
Age	1.29	0.98	0.49	0.16
LF 0.25-0.5	-0.03	0.41	-0.09	0.79
HF1-3	-0.30	0.78	-0.07	0.82
$R^2$ (Adjusted $R^2$ )	0.30(0.07)			
F0 - 9 semitones High target				
Age	-0.39	0.46	-0.31	0.41
LF 0.25-0.5	-0.26	0.37	-0.26	0.51
HF1-3	0.03	0.18	0.05	0.88
$R^2$ (Adjusted $R^2$ )	0.09(-0.22)			

dependent on the sentence materials used. Based on the findings of the present study, it can be concluded that performance in the presence of competing speech with limited linguistic cues was limited primarily by informational masking, even for listeners with hearing loss.

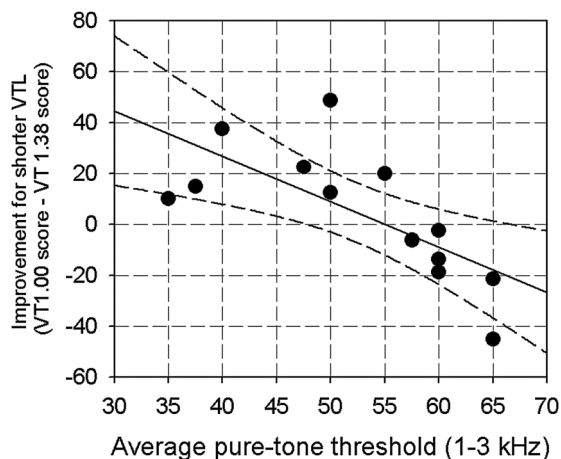


FIG. 8. Improvement in recognition for listeners with hearing loss for a VTL ratio of 1.38 for blocks in which the target was assigned to the shorter VTL, plotted as a function of the average pure-tone threshold between 1 and 3 kHz. The solid line shows a linear regression function fit to the data and the dotted lines define the 95% confidence interval of the regression function.

## B. Use of F0 cues

For normal-hearing listeners, the magnitude of improvement for the higher F0 target was 36 percentage points. This improvement was within the range of improvement reported by Darwin *et al.* (2003) for a talker difference of nine semitones at similar TCRs.

Listeners with hearing loss were able to use talker differences in F0 to improve their recognition of a sentence in the presence of a single competing sentence, but only when the target had the higher F0. For the higher F0 target, the improvement was similar for listeners with and without hearing loss for F0 differences up to six semitones. In contrast, Summers and Leek (1998) found that performance of listeners with hearing loss plateaued at an F0 difference of only two semitones. Audibility and the bandwidth available to the listeners may have contributed to differences between the two studies. Summers and Leek presented stimuli at an overall level of 90 dB sound pressure level without frequency response shaping, which may not have provided sufficient audibility across as wide a range of frequencies as here, especially for listeners with sloping hearing losses. In the current study, stimuli were individually amplified to simulate the response of a hearing aid. Although this does not ensure audibility across the frequency range, it may have provided better high-frequency audibility than a flat response. Other deficits associated with cochlear hearing loss, such as impaired frequency resolution or limited access to temporal fine structure (TFS) may account for the performance plateau observed beyond the six-semitone difference. Because TFS conveys F0 information, the reduction in the ability to use TFS cues (Hopkins *et al.*, 2008; Moore and Moore, 2003) may limit the use of F0 for talker segregation. Also, reduced audibility and frequency resolution at higher frequencies may reduce the usefulness of broad-band cues shown to be involved in the segregation of speech for larger F0 differences (Bird and Darwin, 1998; Culling and Darwin, 1993; Rossi-Katz and Arehart, 2005).

## C. Use of vocal-tract length cues

For listeners with normal hearing, the benefit from a 38% difference in VTL was 24 percentage points averaged across the two lowest TCRs (Experiment 1). This is similar to the effect (approximately 20 percentage points) observed by Darwin *et al.* (2003) for the same TCRs. Listeners with hearing loss did not, on average, benefit from VTL cues, but substantial individual differences were observed. The ability to effectively follow the talker whose formants were shifted to higher values decreased with increasing hearing loss between 1 and 3 kHz. The 1–3 kHz region corresponds to the general range of the second formant. Therefore, a reduction in access to this frequency range may affect the use of formant cues contributing to talker segregation.

Results suggest that even with frequency shaping typical of that used in hearing aids, listeners with 1–3 kHz hearing loss greater than 55 dB hearing level may have difficulty using vocal tract cues for talker segregation. This pattern is consistent with other studies showing that, in general, high-frequency audibility contributes less to speech intelligibility



when high-frequency thresholds exceed 55 dB hearing level (Ching *et al.*, 1998; Hogan and Turner, 1998).

#### D. High–low target asymmetry and uncertainty effects

An asymmetry in scores for the higher and lower target values was observed for talker differences in F0, but not VTL. In contrast to the present study, Darwin *et al.* (2003) did not report asymmetries for F0, but did report them for VTL.

The asymmetry in scores for the higher and lower target F0 is consistent with the findings of Assmann and his colleagues (Assmann, 1999; Assmann and Paschall, 1998), but contrasts with the findings of other investigators (Bird and Darwin, 1998; Darwin *et al.*, 2003; Summers and Leek, 1998; Vestergaard *et al.*, 2009). In the present study, the asymmetry was reduced substantially by the elimination of trial-to-trial target uncertainty, suggesting that attentional factors may be partially responsible. After the removal of uncertainty, however, the asymmetry between scores for high and low target values was still present for both listener groups (15 and 23 percentage points for the groups with normal hearing and hearing loss, respectively). There were, however, several listeners from each group who performed equally well when targets were assigned to the high and low F0. More general effects of target–talker uncertainty have been reported by Brungart and his colleagues (Brungart *et al.*, 2001). They found lower recognition scores when the target talker was randomly selected on each trial than when the target talker was fixed over a block of trials. The effects were minimal, however, when there were only two talkers and the target and competing talkers had the same gender.

The persistence of the asymmetry between high and low target scores in the present study suggests that factors in addition to attention were involved. The possibility that the signal processing used here contributed to the differences between scores for the high and low F0 targets cannot be ignored. In several studies in which F0 asymmetries were not reported (Ives *et al.*, 2010; Vestergaard *et al.*, 2009), the STRAIGHT vocoder (Kawahara *et al.*, 1999) was used. In the current study, the Praat vocoder with PSOLA re-synthesis was used. Darwin *et al.* (2003), however, also used the Praat vocoder and reported no asymmetries for conditions involving F0 manipulations. Relative to the original stimuli, we did observe a slight variable delay (0 to 1 ms; mean 0.44 ms) for stimuli processed with Praat to produce no change in VTL and F0, but this overall delay would not be expected to affect the results. Apart from this delay, the waveforms were the same. Additional distortion specific to the amount of manipulation may be expected to result in inaccuracies in F0 changes. Further analyses of the sentences, however, confirmed that the Praat processing shifted the average F0s as expected. Therefore, it seems unlikely that processing was responsible for the substantial asymmetry observed.

An alternative explanation for the observed F0 asymmetry is increased perceptual salience of the sentence with the higher pitch. It is plausible that the listeners had difficulty “hearing out” the lower F0 target in the same way that listeners have difficulty “hearing out” lower pitch components of

polyphonic music (Gregory, 1990; Palmer and Holleran, 1994). Palmer and Holleran (1994), for example, reported that detection of pitch changes in three-part polyphonic music is less noticeable for lower frequency melodies in the presence of higher frequency melodies. Similarly, Gregory (1990) demonstrated that recognition of melodies in polyphonic music was significantly poorer for the low-frequency melodies (55%) than for the high frequency melodies (91%), even when the melody lines for the higher and lower registers were swapped. The explanation in terms of pitch salience is also consistent with the observation that the pitch of two concurrent vowels is dominated by the higher of the two F0 frequencies (Assmann and Paschall, 1998).

The above arguments, though plausible, do not fully explain why F0 target asymmetries have not been consistently reported. A possible factor that may account for differences among the studies is listener age. Listeners in the present study were generally older than for studies in which asymmetries were not reported. Age-related differences in listeners’ abilities to selectively attend to a less dominant pitch may have contributed to the differences between the present findings and those of several other studies.

On average, there was no asymmetry between responses to the shorter or longer VTLs in the present study for either listener group. This finding is consistent with results reported by Vestergaard *et al.* (2009). Darwin *et al.* (2003), however, reported that normal-hearing listeners had higher scores for sentences with the shorter vocal tract length. In addition to the fact that listener characteristics differed among the studies, only one talker was used in the present study and in the study by Vestergaard *et al.* (2009), whereas the findings of Darwin *et al.* were based on several talkers.

#### E. Listener age

Age was not found to affect the use of talker segregation cues by listeners with hearing loss in the present study. This finding contrasts with results of other studies (Humes and Coughlin, 2009; Rossi-Katz and Arehart, 2009; Summers and Leek, 1998). Greater hearing loss is often coupled with increased listener age; however, making it difficult to completely separate the influences of these factors. Unlike other studies, the older listeners with hearing loss in the present study tended to have less hearing loss than the younger listeners, which may explain the absence of age as a contributing factor.

#### F. Generalizability

There are several aspects of the current study that affect the generalizability of the findings. First, the speech CRM stimuli contain a limited vocabulary and minimal contextual cues. While the CRM corpus has the advantage of enabling better isolation of the acoustic cues influencing talker segregation than more contextually rich materials, the interactions between acoustic and contextual cues and their relative importance cannot be examined.

In the current study, the F0 and spectral envelopes of the sentence pairs were manipulated independently. Although independent manipulation of cues has experimental advantages, these manipulations contrast with normal



speech in which average talker differences in F0 and VTL would co-vary. In addition, the extent to which listeners with hearing loss can use F0 and vocal-tract length cues in multi-talker or more spectro-temporally dense backgrounds is yet to be determined.

## G. Implications

For listeners with moderate hearing loss with appropriate amplification, the benefit of F0 cues for talker segregation was similar to that for normal-hearing listeners for all but the largest F0 difference. The lack of benefit from VTL cues, however, indicates that listeners with hearing loss may be unable to take advantage of talker and gender differences in VTL. A possible avenue for further research is the use of signal processing techniques that enhance VTL differences and make them more accessible to listeners with hearing loss. There may also be options for source-separation algorithms that provide the listeners with pre-separated signals (Lui *et al.*, 2008; Wang and Brown, 2006; Wu and Wang, 2006).

## VII. CONCLUSIONS

1. Identification of CRM sentences in the presence of AM noise was substantially better than in the presence of competing sentences, suggesting that, even for listeners with hearing loss, informational, rather than energetic masking, was the dominant source of interference.
2. Like normally hearing adults, adults with moderate hearing loss and appropriate amplification could use differences in average F0 to separate competing talkers in the absence of VTL cues, for F0 differences up to six semitones.
3. Under conditions of trial-to-trial target uncertainty, however, this effect was only evident when the target talker had the higher F0.
4. Even when trial-to-trial uncertainty was removed, recognition was better for the talker with the higher F0.
5. On average, unlike normally hearing listeners, the ability of adults with moderate hearing loss to separate competing talkers on the basis of VTL differences in the absence of differences of average F0 was small or non-existent.
6. To the extent that some individuals with moderate hearing loss may be able to use VTL, the effects appear to be associated with reasonably good thresholds (< 60 dB hearing level) in the region of the second formant.

## ACKNOWLEDGMENTS

This work was funded by the National Institute of Deafness and Communicative Disorders (NIDCD DC007500-01A2). We gratefully acknowledge the valuable comments of Brian Moore, Marjorie Leek, and one anonymous reviewer on earlier versions of the manuscript.

ANSI S3.5 (1997). "American National Standard Methods for the Calculation of the Speech Intelligibility Index," American National Standards Institute, New York.

Arehart, K. H. (1994). "Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners," *J. Acoust. Soc. Am.* **95**, 3574–3585.

Arehart, K. H. (1998). "Effects of high-frequency amplification on double-vowel identification in listeners with hearing loss," *J. Acoust. Soc. Am.* **104**, 1733–1736.

Arehart, K. H., King, C. A., and McLean-Mudgett, K. S. (1997). "Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss," *J. Speech Lang. Hear. Res.* **40**, 1434–1444.

Arehart, K. H., Rossi-Katz, J., and Swensson-Prutsmann, J. (2005). "Double-vowel perception in listeners with cochlear hearing loss: differences in fundamental frequency, ear of presentation, and relative amplitude," *J. Speech Lang. Hear. Res.* **48**, 236–252.

Assmann, P. F. (1999). "Fundamental frequency and the intelligibility of competing sentences," in Abstract A290, Midwinter Meeting of the Association of Research in Otolaryngology, St. Petersburg Beach, FL, available at <http://www.aro.org/archives/1999/290.html> (Last viewed 7/22/2011).

Assmann, P. F., and Paschall, D. D. (1998). "Pitches of concurrent vowels," *J. Acoust. Soc. Am.* **103**, 1150–1160.

Assmann, P. F., and Summerfield, Q. (2004). "Perception of speech under adverse conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 231–308.

Bernstein, J. A., and Oxenham, A. J. (2006). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," *J. Acoust. Soc. Am.* **120**, 3929–3945.

Bird, J., and Darwin, C. J. (1998). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr Publishers, Grantham, U.K.), pp. 263–269.

Boersma, P., and Weenink, D. (2006). "Praat: Doing phonetics by computer (version 4.1.2.1) [computer program]," <http://www.praat.org/> (last viewed July 5, 2006).

Bolia, R., Nelson, W., Ericson, M., and Simpson, B. (2000). "A speech corpus for multitalker communications research," *J. Acoust. Soc. Am.* **107**, 1065–1066.

Bregman, A. S. (1990). *Auditory Scene Analysis* (MIT Press, Cambridge, MA), pp. 529–554.

Brokx, J. P. L., and Nootboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," *J. Phon.* **10**, 23–36.

Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* **109**, 1101–1109.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (2006). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," *J. Acoust. Soc. Am.* **120**, 4007–4018.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," *J. Acoust. Soc. Am.* **110**, 2527–2538.

Byrne, C., and Cotton, S. (1988). "Evaluation of the National Acoustic Laboratories' new hearing aid selection procedure," *J. Speech Hear. Res.* **31**, 178–186.

Ching, T. Y., Dillon, H., and Byrne, D. (1998). "Speech recognition of hearing-impaired listeners predictions from audibility and the limited role of high-frequency amplification," *J. Acoust. Soc. Am.* **103**, 1128–1140.

Culling, J. F., and Darwin, C. J. (1993). "Perceptual separation of simultaneous vowels: Within and across formant grouping by Fo," *J. Acoust. Soc. Am.* **93**, 3454–3467.

Dai, H. (2000). "On the relative influence of individual harmonics on pitch judgment," *J. Acoust. Soc. Am.* **107**, 953–959.

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," *J. Acoust. Soc. Am.* **114**, 2913–2922.

Darwin, C. J., and Hukin, R. W. (2000a). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," *J. Acoust. Soc. Am.* **107**, 970–977.

Darwin, C. J., and Hukin, R. W. (2000b). "Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention," *J. Acoust. Soc. Am.* **108**, 335–342.

Drullman, R., and Bronkhorst, A. W. (2004). "Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers," *J. Acoust. Soc. Am.* **116**, 3090–3098.

Durlach, N., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). "Note on informational masking," *J. Acoust. Soc. Am.* **113**, 2984–2987.

- Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state," A practical method for grading the cognitive state of patients for the clinician," *J. Psychiatr. Res.* **12**, 189–198.
- Gregory, A. H. (1990). "Listening to polyphonic music," *Psychol. Music* **18**, 163–170.
- Grimault, N., Michey, C., Carlyon, R., and Arthaud, L. (2000). "Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency," *J. Acoust. Soc. Am.* **108**, 263–271.
- Grose, J. H., and Hall, J. (1996). "Perceptual organization of sequential stimuli in listeners with cochlear hearing loss," *J. Speech Lang. Hear. Res.* **39**, 1149–1158.
- Hogan, C., and Turner, C. (1998). "High frequency audibility: Benefits for hearing-impaired listeners," *J. Acoust. Soc. Am.* **104**, 432–441.
- Hopkins, K., Moore, B. C. J., and Stone, M. A. (2008). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," *J. Acoust. Soc. Am.* **123**, 1140–1153.
- Humes, L. E., and Coughlin, M. (2009). "Aided speech-identification performance in single-talker competition by older adults with impaired hearing," *Scand. J. Psychol.* **50**, 485–494.
- Ives, D. T., Smith, D. R. R., and Patterson, R. D. (2005). "Discrimination of speaker size from syllable phrases," *J. Acoust. Soc. Am.* **118**, 3816–3822.
- Ives, D. T., Vestergaard, M. D., Kistler, D. J., and Patterson, R. D. (2010). "Location and acoustic scale cues in concurrent speech recognition," *J. Acoust. Soc. Am.* **127**, 3729–3737.
- Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. (1999). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction," *Speech Commun.* **27**, 187–207.
- Leek, M. R., and Summers, V. (2001). "Pitch strength and pitch dominance of iterated rippled noises in hearing-impaired listeners," *J. Acoust. Soc. Am.* **109**, 2944–2954.
- Lui, J., Xin, J., and Qi, Y. (2008). "A dynamic algorithm for separation of convolutive sound mixtures," *Neuro-computing* **72**, 521–532.
- Mackersie, C. L., Prida, T., and Stiles, D. (2001). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss," *J. Speech, Lang. Hear. Res.* **44**, 19–28.
- Moore, B. C. J., Glasberg, B. R., and Peters, R. (1985). "Relative dominance of individual partials in determining the pitch of complex tones," *J. Acoust. Soc. Am.* **77**, 1853–1860.
- Moore, B. C. J., and Moore, G. A. (2003). "Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects," *Hear. Res.* **182**, 153–163.
- Moore, B. C. J., and Peters, R. W. (1992). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," *J. Acoust. Soc. Am.* **91**, 2881–2893.
- Moulines, E., and Charpentier, F. (1990). "Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Commun.* **9**, 453–467.
- Palmer, C., and Holleran, S. (1994). "Harmonic, melodic, and frequency height influences in the perception of multivoiced music," *Percept. Psychophys.* **56**, 301–312.
- Peters, R. W., Moore, B. C. J., and Baer, T. (1998). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," *J. Acoust. Soc. Am.* **103**, 577–587.
- Rose, M. M., and Moore, B. C. J. (1997). "Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **102**, 1768–1778.
- Rose, M. M., and Moore, B. C. J. (2000). "Effects of frequency and level on auditory stream segregation," *J. Acoust. Soc. Am.* **108**, 1209–1214.
- Rossi-Katz, J. A., and Arehart, K. H. (2005). "Effects of cochlear hearing loss on perceptual grouping cues in competing-vowel perception," *J. Acoust. Soc. Am.* **118**, 2588–2598.
- Rossi-Katz, J. A., and Arehart, K. H. (2009). "Message and talker identification in older adults: effects of task, distinctiveness of the talkers' voices, and meaningfulness of the competing message," *J. Speech Lang. Hear. Res.* **52**, 435–453.
- Schneider, B. A., Li, L., and Daneman, M. (2007). "How competing speech interferes with speech comprehension in everyday listening situations," *J. Am. Acad. Audiol.* **18**, 559–572.
- Silman, S., and Gelfand, S. (1981). "The relationship between magnitude of hearing loss and acoustic reflex thresholds," *J. Speech Hear. Disord.* **46**, 312–316.
- Smith, D. R. R., and Patterson, R. D. (2005). "The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age," *J. Acoust. Soc. Am.* **118**, 3177–3186.
- Studebaker, G. A. (1985). "A 'Rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Summers, V., and Leek, M. R. (1998). "F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss," *J. Speech Lang. Hear. Res.* **41**, 1294–1306.
- Vestergaard, M. D., Fyson, N. R. C., and Patterson, R. D. (2009). "The interaction of vocal characteristics and audibility in the recognition of concurrent syllables," *J. Acoust. Soc. Am.* **125**, 1114–1124.
- Vestergaard, M. D., and Patterson, R. D. (2009). "Effects of voicing in the recognition of concurrent syllables," *J. Acoust. Soc. Am.* **126**, 2860–2863.
- Wang, D. L., and Brown, G. J. (2006). "Fundamentals of computational auditory scene analysis," in *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, edited by D. L. Wang and G. J. Brown (Wiley, Hoboken, NJ), pp. 1–44.
- Watson, C. S. (2005). "Some comments on informational masking," *Acta Acust.* **91**, 502–512.
- Wu, M., and Wang, D. (2006). "A two-stage algorithm for one-microphone reverberant speech enhancement," *IEEE Trans. Audio. Speech Lang. Processing* **14**, 774–784.