Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background[a)]

Jae Hee Lee[b)] and Larry E. Humes
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

This study investigated the benefits of differences between sentences in fundamental frequency ($F0$) and temporal onset for sentence pairs among listener groups differing in age and hearing sensitivity. Two experiments were completed with the primary difference between experiments being the way in which the stimuli were presented. Experiment 1 used blocked stimulus presentation, which ultimately provided redundant acoustic cues to mark the target sentence in each pair, whereas Experiment 2 sampled a slightly more restricted stimulus space, but in a completely randomized presentation order. For both experiments, listeners were required to detect a cue word ("Baron") for the target sentence in each pair and to then identify the target words (color, number) that appeared later in the target sentence. Results of Experiment 1 showed that $F0$ or onset separation cues were beneficial to both cue-word detection and color-number identification performance. There were no significant differences across groups in the ability to detect the cue word, but groups differed in their ability to identify the correct color-number words. Elderly adults with impaired hearing had the greatest difficulty with the identification task despite the application of spectral shaping to restore the audibility of the speech stimuli. For the most part, the primary results of Experiment 1 were replicated in Experiment 2, although, in the latter experiment, all older adults, whether they had normal or impaired hearing, performed worse than young adults with normal hearing. From Experiment 2, the benefits received for a difference in $F0$ between talkers of 6 semitones were equivalent to those received for an onset asynchrony of 300 ms between sentences and, for such conditions, the combination of both sound-segregation cues resulted in an additive benefit.
© 2012 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4740482]

I. INTRODUCTION

For multi-talker conversations in which the target and the interfering messages are presented concurrently, listeners are required to encode and contrast spectral and temporal features of both competing speech signals, segregate the target from the competing source, and use a cognitive strategy to selectively attend to the target message while inhibiting the competing information. The processes involved in segregating competing speech signals may be peripheral, central-auditory, or cognitive in nature [see Shinn–Cunningham and Best (2008) for a review]. Age-related changes can negatively affect both the hearing sensitivity responsible for the peripheral encoding of input signals and the processing of speech by the central-auditory or cognitive systems (CHABA, 1988) and, as a result, age-related deficits in multi-talker speech perception have been documented in numerous studies (Helfer et al., 2010; Helfer and Freyman, 2008; Humes and Coughlin, 2009; Humes et al., 2006; Rossi–Katz and Arehart, 2009; Tun et al., 2002; Tun and Wingfield, 1999).

To improve intelligibility in multi-talker speech communication, acoustical features of the competing speech messages, such as fundamental frequency and corresponding harmonics, onset time, intensity, frequency, total duration, and spatial location, can facilitate the segregation of competing speech signals. Differences in fundamental frequency ($\Delta F0$) and onset are two different and potentially powerful temporal segregation cues when processing monaurally presented competing speech signals (Bregman, 1990; Darwin, 2001; Hedrick and Madix, 2009; Lentz and Marsh, 2006). Information from $\Delta F0$ is mainly carried by the temporal fine structure and this information is preserved in the temporal pattern of auditory nerve firings. The temporal onset asynchrony between two signals is an important gross temporal cue extracted by the temporal onset disparity in the envelopes of two sounds. Because two different speakers usually have different $F0$ values and often do not start speaking at the same time both $F0$ and onset segregation cues are often likely to occur in our everyday listening. As a result, it is important to examine the role of $\Delta F0$ and onset asynchrony between two competing speech signals and any negative impact of aging on the use of these two sound-segregation cues.

A. Perceptual benefit from F0 differences between competing speech signals

The fundamental frequency, $F0$, is defined as the frequency at which the vocal folds vibrate when voiced speech

---

[b)]Author to whom correspondence should be addressed. Current address: Department of Audiology, Hallym University of Graduate Studies, 906-18 Daechidong Gangnamgu, Seoul 135-841, Republic of Korea. Electronic mail: leejaehee@hallym.ac.kr

sounds are made. The $F0$ value conveys various cues at subsegmental, segmental, and suprasegmental levels, such as acoustic cues to vowel identity (different $F0$ and formant frequencies across vowels), gender (lower $F0$ in men), age (a decrease in mean $F0$ with age, more pronounced in women), intonation (greater $F0$ fluctuations for a greater change in intonation), and the speaker's emotional state (higher $F0$ in "happy" or lower $F0$ in "sad" emotional state) (e.g., Cooper and Sorensen, 1981; Gelfer and Mikos, 2005; Harrington et al., 2007; Murray and Arnott, 1993; Peterson and Barney, 1952).

A review of the literature on the effects of $\Delta F0$ can be summarized briefly as follows. First, $\Delta F0$ significantly benefited identification performance whether the competing speech signals are steady-state synthesized vowel pairs (Alain et al., 2005; Arehart et al.et al., 1997, 2005; Assmann and Summerfield, 1990, 1994; Chalikia and Bregman, 1989; Culling and Darwin, 1993, 1994; de Cheveigné, 1997; Meddis and Hewitt, 1992; Rossi–Katz and Arehart, 2005; Stubbs and Summerfield, 1988; Summerfield and Assmann, 1989, 1991; Summers and Leek, 1998; Vongpaisal and Pichora–Fuller, 2007), nonsense syllables (Vestergaard et al., 2009), sentence pairs without natural $F0$ variation (Assmann, 1999; Bird and Darwin, 1998; Brokx and Nooteboom, 1982), or with natural $F0$ variation preserved (Assmann, 1999; Darwin et al., 2003; Oxenham and Simonson, 2009; Summers and Leek, 1998). However, the pattern of improvement was observed to be gradual over a greater range of $\Delta F0$s [up to $\Delta F0$ of 8 or 9 semitones (STs)] for sentence-identification whereas the vowel-identification performance at $\Delta F0$ of ≥ 2 ST reached asymptote apparently due to the pattern of beating between double vowels (Culling and Darwin, 1994). These findings suggest a lack of a strong relation between double-vowel and double-sentence identification performance, partially related to the ceiling performance in double-vowel paradigm.

Second, listeners with impaired hearing received perceptual benefits for $\Delta F0$, but often showed less of a $\Delta F0$ benefit in the double-vowel and double-sentence tasks compared to normal-hearing (NH) listeners (Arehart, 1998; Arehart et al., 1997, 2005; Mackersie et al., 2011; Rossi–Katz and Arehart, 2005; Stubbs and Summerfield, 1988; Summers and Leek, 1998). Summers and Leek (1998) found less of a benefit from $\Delta F0$ (4 ST) in some hearing-impaired (HI) participants, and reported a significant relation between the $\Delta F0$ benefits and the high-frequency hearing thresholds of individuals. Rossi–Katz and Arehart (2005) showed that cochlear hearing loss negatively influenced listeners' ability to use $\Delta F0$ for within-formant grouping in the high-frequency region. However, Arehart (1998) found no additional $\Delta F0$ benefit to vowel-identification performance when compensating for the reduced audibility of vowels at high frequencies. Mackersie et al. (2011) also reported less of an improvement from 9 ST of $\Delta F0$ (high-$F0$ target) in HI listeners (age range: 45 to 76 yr) than in NH listeners (age range: 25 to 69 yr), even with individually amplified stimuli. However, high-frequency hearing thresholds from 1 to 3 kHz could not predict the magnitude of benefit from 9 ST of $\Delta F0$.

Third, age effects have been examined less frequently but there is increasing evidence that an age-related loss of neural synchrony may occur, increasing temporal jitter and degrading benefit from $F0$ differences for concurrent speech signals (Grose and Mamo, 2010; Vongpaisal and Pichora–Fuller, 2007). Age-related reduction in the synchrony of auditory nerve fiber responses has also been found in various animal studies (Backoff and Caspary, 1994; Boettcher et al., 1996; Mills et al., 2006; Raza et al., 1994). Previous behavioral studies have found that older individuals with NH were less accurate at identifying concurrent vowels as a function of $\Delta F0$, with 0–4 ST between vowel pairs, compared to the young normal hearers (Arehart et al., 2011; Summers and Leek, 1998; Vongpaisal and Pichora–Fuller, 2007).

In order to examine any negative effects of hearing loss or other age-related declines on the use of $\Delta F0$, the current study carried out between-group comparisons among four listener groups differing in hearing status and age. Although the $\Delta F0$ cue per se is a low-level physical segregation cue encoded in the periphery, the processing of the $F0$-guided competing speech signals needs to go through all the peripheral, central, and cognitive mechanisms. The present study compensated for the reduced audibility of the elderly hearing-impaired (EHI) listeners by introducing amplitude adjustments to the spectra of speech materials, similar to the role of clinical amplification. If age-related processing deficits exist in the ability to use $\Delta F0$, regardless of the compensation for age-related peripheral hearing loss, we would still expect to observe a diminished $\Delta F0$ benefit. Also, young adults with either NH or hearing loss simulated by masking noise allow a comparative evaluation of the roles of peripheral inaudibility and higher level processing with regard to $\Delta F0$ benefit.

B. Perceptual benefit from onset asynchrony between competing speech signals

Onset asynchrony between competing signals is one of the powerful sound segregation cues, especially when $F0$, amplitude, or spatial cues between competing signals are unavailable (Bregman, 1990; Darwin, 1981, 2001, 2008). The ability of listeners to use onset asynchrony between speech signals is undoubtedly important since multiple talkers are unlikely to speak with exactly the same onset in a realistic communication situation.

The advantage from temporal onset asynchrony between competing speech signals has been investigated in only a few studies using synthesized competing vowels (Hedrick and Madix, 2009; Lentz and Marsh, 2006; Summerfield and Assmann, 1989; Summerfield and Culling, 1992). No studies of an onset asynchrony benefit have been conducted with competing sentences.

Using two synthesized vowels with durations of 400 ms, Summerfield and Culling (1992) measured the minimal signal-to-noise ratio at which listeners could just identify a target vowel against a masking vowel. When the double vowels have the same $F0$, but the masker vowel started 200 ms before the target vowel, the 200-ms onset asynchrony significantly reduced (improved) masked threshold by approximately 5–6 dB.

Summerfield and Assmann (1989) introduced a 1-s precursor, specifying one of two subsequent concurrent vowels within the precursor, presented either auditorily (ipsilaterally or contralaterally) or visually. Only the ipsilateral acoustic precursor significantly improved vowel-identification accuracy compared to the conditions without a precursor, suggesting a similar positive role of either the precursor or the leading segment corresponding to the asynchrony.

Hedrick and Madix (2009) investigated the ability of young normal hearers in identifying synthetic double vowels with various onset separations from 0 to 150 ms (in 25-ms steps). Consistent with other studies, a robust onset-separation benefit was observed when vowel identification was measured with the same-$F0$ vowels. Despite the use of relatively difficult vowel pairs in their study, some vowel pairs revealed a significant onset effect or a vowel dominance pattern while others did not. Moreover, the harmonic structures of vowels could not account for the differential vowel dominance. Considering the perceptual onset benefit and the vowel dominance pattern together, the authors concluded that the process of onset asynchrony between double vowels would involve a schema-based categorical process beyond the auditory peripheral level, at least for the young normal-hearing (YNH) population.

Besides data from normal hearers noted above, Lentz and Marsh (2006) measured the ability of both NH (age range: 18–51, $M = 31$ yrs) and HI participants (age range: 25–61, $M = 45.5$ yrs) to utilize onset asynchrony as a segregation cue.

The synthesized competing vowels were separated by various onset asynchronies but shared the same offset. In their double-vowel paradigm, the masker vowel, which came first, had different durations from 350 to 550 ms while the target vowel, which always came later, had a duration fixed at 250 ms. As a result, the duration of temporal overlap was always 250 ms, with onset asynchrony ranging from 100 to 300 ms. Both groups' identification accuracy of the lagging target vowel continued to improve as a function of onset separations from 100 to 300 ms, regardless of whether the double vowels were of the same or different $F0$. Between groups, the HI group received a slightly less perceptual benefit from 100–300 ms onset separations relative to the NH group (13 percentage points and 9 percentage points of perceptual benefit in the NH and HI groups, respectively). In contrast, the $\Delta F0$ cue (4 ST) yielded a similar benefit in both groups (about 11 percentage points).

Given that the amount of onset benefit was slightly smaller in the HI group than in the NH group and that spectral shaping was not applied to the vowels to compensate for hearing loss, Lentz and Marsh (2006) further examined the relation between the audibility of the speech and the onset asynchrony benefit using an excitation-pattern analysis. The results showed that both the excitation pattern and the audible speech frequency range failed to predict the onset separation benefits of HI individuals, arguing that the use of onset asynchrony would be associated with the presence of suprathreshold deficits similar to the suprathreshold deficits suggested for diminished $\Delta F0$ benefit (Arehart, 1998).

It is somewhat surprising that no studies have yet investigated the perceptual benefit of onset asynchrony between competing messages for sentences and the effect of aging on such benefit. Although both $\Delta F0$ and onset asynchronies are often occurring in common multi-talker listening environments and aging could potentially diminish the benefits from each cue, previous research has seldom focused on the $F0$-guided or onset-guided segregation strategy of the old listeners for the concurrent messages. Clearly, some form of temporal comparison between stimuli would be required to make use of onset asynchrony.

C. Purpose of the study

This study was designed to compare the performance of the four groups of adults, differing in age and hearing sensitivity, when they were asked to detect and identify double sentences separated by $\Delta F0$, onset asynchrony, or a combination of each. Specifically, as described in more detail below, the four groups were YNH adults, elderly normal-hearing (ENH) adults, EHI adults, and YNH adults with noise masking (YNM) designed to create the same inaudibility as that experienced by the EHI listeners.

The present study goes beyond previous research in three ways. First, the present study systematically manipulated differences in both $\Delta F0$ and temporal onset asynchrony, allowing us to compare the $\Delta F0$ and onset benefits across the four groups, as well as to evaluate any interactive effects between the $F0$ and onset separations in each group. By comparing the performance across various pairs of the four groups included in this study, the relative importance of inaudibility, peripheral cochlear pathology, and aging was examined. For example, comparisons between the performance of YNH and YNM groups enable the evaluation of a relative role of a sensitivity loss on performance. Similarly, a comparison of ENH and EHI allows an examination of sensitivity loss and cochlear pathology on performance. In a similar fashion, a comparison of the performance of the YNH and ENH groups presents an examination of the effects of age on performance. Finally, the pattern of group results for all these experimental groups relative to the YNH reference group may also shed light on the interactions between inaudibility/pathology and age.

Second, using the coordinate response measure (CRM) corpus (Bolia et al., 2000), a stimulus corpus frequently used for investigating competing speech perception, we measured both the speech-detection and identification performance when two sound-segregation cues ($F0$ and onset differences) were available to the listeners. Although various studies have measured the performance of young (Brungart et al., 2006; Brungart and Simpson, 2007) and older listeners (Humes and Coughlin, 2009, Humes et al., 2006) on competing-speech tasks using the CRM, the focus of those studies was mostly on the identification of target color-number (CN) words near the end of each sentence. In a multi-talker conversation simulated by competing CRM sentences, the target CN words have most often been cued by the cue-word "Baron" appearing early in the target sentence. Thus, to correctly identify the CN words in the target

sentence, the listeners must first be able to detect the cue word identifying the target message. As a result, if one subject group performs worse than another in terms of correct CN identification, there are at least two plausible explanations. First, difficulty in detecting the cue word could result in poor identification of color and number words spoken by the target voice. Second, listeners may be able to identify the correct cue word and know which sentence is the target sentence but may still not be able to identify the correct CN words that appear later in the mixture of target and competing sentences. To sort these two possibilities out, Shafiro and Gygi (2007) first measured detection sensitivity (d') to the cue word Baron in YNH listeners and found that d' was strongly correlated with the identification accuracy of the CN words in the target message. As noted, age-group differences in CN identification have been observed previously (Humes and Coughlin, 2009; Humes et al., 2006) but it is unclear whether such performance decrements are due to difficulty in identifying the initial cue word marking the target message. As a result, this issue is explored further in this study.

Third, two experiments were conducted in the present study in order to determine the effect of uncertainty on the use of $F0$ and onset segregation cues for detection and identification performance. This was motivated given the contradictory findings on the contribution of uncertainty to speech recognition between young and old listener groups (Brungart and Simpson, 2004; Freyman et al., 2007; Humes and Coughlin, 2009; Humes et al., 2006; Mackersie et al., 2011; Sommers, 1997). For the competing speech environment, it is well known that target-masker similarity and stimulus uncertainty are associated with informational masking (Durlach et al., 2003; Freyman et al., 2007).

Freyman et al. (2007) varied the amount of masker uncertainty for a speech-recognition task and expected to observe more informational masking with increasing uncertainty of the masker. Unexpectedly, they found a relatively small effect of masker uncertainty on nonsense sentence identification for YNH listeners. Brungart and Simpson (2004), using the CRM corpus, also found very little effect of talker uncertainty on speech-identification for young adults. However, Mackersie et al. (2011), also using the CRM, found that, when the trial- to-trial target uncertainty was removed, both NH (mean age: 48 yr, range: 25–69 yr) and generally older HI listeners (mean age: 61 yr, range: 45–76 yr) better identified the target CN words, especially for the target talker having a lower $F0$ value. This suggested that the target uncertainty affected listeners' focusing on the lower pitch of the target voice in the presence of the higher-$F0$ competition, regardless of hearing status. Similarly, Humes and Coughlin (2009) and Humes et al. (2006), both using the CRM, observed a significant effect of talker uncertainty in both young and older adults.

In the current study, Experiment 1 examined the contributions of $F0$ and onset segregation cues using finer steps along each cue continuum and using minimum or low-uncertainty test conditions. Experiment 2 was essentially a replication of much of Experiment 1, but with completely randomized (maximum uncertainty) stimulus conditions.

D. Hypotheses tested

Based on the literature reviewed above, the following hypotheses have been developed and will be evaluated in the first experiment: (1) $F0$ differences will improve performance of the four listener groups on the detection and identification of the target words, and the improvement will be gradual from 0 to 6 ST rather than asymptotic; (2) onset asynchrony will also significantly enhance the ability of all the groups to detect and identify the target words, with performance progressively increasing from 0 to 600 ms asynchrony; (3) combinations of $\Delta F0$ and onset asynchrony cues will result in a higher performance than either $\Delta F0$ or onset-asynchrony alone; (4) the overall identification performance of the EHI group will be significantly poorer than that of the YNH group, despite audible sentence pairs, with the performance of the other groups (YNM and ENH) somewhere in between these two groups; and (5) the ability of the EHI group to use $\Delta F0$ and onset asynchrony cues will be poorer than that of the YNH group. The methods used to examine these hypotheses follow.

II. EXPERIMENT 1: METHODS

A. Participants

Sixty listeners, consisting of 4 groups of 15 listeners, participated in the first experiment. The four groups were as follows: (1) 15 YNH adults with ages ranging from 19 to 32 yr ($M = 23.5$, SE = 1 yrs); (2) 15 ENH adults between the ages of 63 and 79 yrs ($M = 70.3$, standard error (SE) = 1.4 yrs); (3) 15 EHI adults between the ages of 61 and 81 yrs ($M = 71.3$, SE = 1.8 yrs); and (4) 15 YNM listeners (YNH listeners with noise masking) to simulate the average audibility of the EHI listeners with ages ranging from 19 to 26 yrs ($M = 21.8$, SE = 0.5 yrs). All the participants were required to have normal middle-ear status (normal tympanogram), a score of at least 25 out of 30 on the Mini-Mental Status Exam (MMSE, Folstein et al., 1975) for cognitive status, and a score of 9 or greater when summed from the auditory forward and backward digit-span test for memory from the Wechsler Adult Intelligence Scale (Wechsler, 1997).

The YNH and YNM listener groups were screened to ensure that their air conduction thresholds were equal to or better than 20 dB hearing level (HL) (ANSI, 2004) at octave frequencies from 250 through 8000 Hz. The mean thresholds of the 15 ENH participants were equal to or better than 20 dB HL at octave frequencies from 250 through 4000 Hz. Table I shows ages, MMSE scores, digit spans, and air-conduction audiometric thresholds of the test ear for the ENH and EHI individuals. Pair-wise independent sample t-tests indicated that the two groups of young adults did not differ in age, the two groups of older adults did not differ in age, but each group of older adults was significantly older than each of the younger groups. Results of a one-way analysis of variance (ANOVA) showed that the four listener groups did not differ in MMSE [$F(3,56) = 1.5$, $p = 0.23$] or digit-span scores [$F(3,56) = 1.8$, $p = 0.16$].

To examine the effect of hearing loss among the elderly listeners, an overlap in the range of air-conduction thresholds

TABLE I. Demographic information (i.e., age, scores of MMSE and digit-span, and air-conducted audiometric thresholds) of individual ENH and EHI listeners (ENH = the elderly normal-hearing; EHI = the elderly hearing-impaired; MMSE = Mini-Mental Status Exam; PTA 1,2,4 or PTA 0.5,1,2 = Pure-tone thresholds averaged across 1, 2, 4 kHz or across 0.5, 1, 2 kHz).

| | | | Digit | | | | Frequency | | | | | | PTA | PTA |
	Age	MMSE	spans	250 Hz	500 Hz	1000 Hz	1500 Hz	2000 Hz	3000 Hz	4000 Hz	6000 Hz	8000 Hz	1,2,4	0.5,1,2
ENH 1	65	29	21	15	15	15	15	20	15	10	15	40	15	17
ENH 2	79	30	18	25	20	15	10	5	10	15	15	35	12	13
ENH 3	72	28	20	10	5	10	10	5	5	10	15	0	8	7
ENH 4	69	30	14	20	25	20	20	20	20	25	25	40	22	22
ENH 5	69	29	17	15	15	10	13	15	15	30	35	25	18	13
ENH 6	66	30	16	10	15	5	15	10	25	35	35	30	17	10
ENH 7	63	30	17	20	20	20	20	20	25	25	35	30	22	20
ENH 8	64	29	18	15	10	10	5	5	10	5	5	5	7	8
ENH 9	69	30	16	20	20	20	10	15	15	10	15	25	15	18
ENH 10	79	29	13	5	5	10	15	15	10	15	15	40	13	10
ENH 11	72	28	17	20	20	15	15	10	15	20	5	5	15	15
ENH 12	67	25	17	15	10	10	5	10	20	30	30	25	17	10
ENH 13	67	30	20	5	5	5	8	10	20	20	10	20	12	7
ENH 14	79	28	17	10	25	20	5	10	20	15	35	15	15	18
ENH 15	75	29	16	0	5	5	10	10	25	30	20	15	15	7
mean	**70**	**29**	**17**	**14**	**14**	**13**	**12**	**12**	**17**	**20**	**21**	**23**	**15**	**13**
EHI 1	61	30	12	20	20	25	35	35	60	65	65	65	42	27
EHI 2	80	29	21	20	25	30	35	40	45	65	65	75	45	32
EHI 3	63	30	22	40	40	40	35	50	65	65	60	65	52	43
EHI 4	67	29	13	20	30	40	42	45	55	50	55	75	45	38
EHI 5	78	29	18	40	35	35	45	55	60	65	55	65	52	42
EHI 6	63	30	14	30	35	45	50	50	50	60	60	60	52	43
EHI 7	75	28	13	20	25	25	40	45	50	55	65	75	42	32
EHI 8	81	29	13	40	35	30	35	40	45	60	70	75	43	35
EHI 9	71	29	19	25	25	30	35	35	45	60	50	60	42	30
EHI 10	81	29	13	20	25	35	45	55	55	60	65	75	50	38
EHI 11	71	29	10	15	25	25	35	45	45	50	50	60	40	32
EHI 12	76	28	14	20	35	45	55	60	50	55	60	70	53	47
EHI 13	68	29	14	20	25	35	40	45	45	55	60	60	45	35
EHI 14	64	29	9	20	30	40	45	50	55	60	65	60	50	40
EHI 15	70	29	16	15	20	45	45	60	65	65	70	70	57	42
mean	**71**	**29**	**15**	**24**	**29**	**35**	**41**	**47**	**53**	**59**	**61**	**67**	**47**	**37**

between the ENH and EHI individuals at frequencies from 1000 to 8000 Hz was avoided. As shown in Table I, individual audiometric thresholds of the ENH subjects ranged from 5–20 dB HL across frequencies of 1000, 1500, and 2000 Hz in contrast to the range of 25–60 dB HL for EHI subjects. The individual audiometric thresholds of the ENH individuals ranged from 5–25 dB HL at 3000 Hz and 0–40 dB HL from 4000–8000 Hz, whereas the corresponding thresholds for the EHI subjects ranged from 45–65 dB HL at 3000 Hz and 50–75 dB HL from 4000–8000 Hz. Given the monaural presentation condition in this study, if hearing was asymmetrical, the ear that matched best with the range of hearing thresholds for the respective group of older adults was selected for testing. If both ears were eligible as a test ear, then the right ear was selected as the test ear. The right ear was the test ear for all YNH and YNM subjects, for nine ENH subjects, and for eight EHI subjects.

A brief survey of their highest level of education was also completed. Responses for the highest grade of education were coded as follows: 12 for the completion of high school,

16 for Bachelor's degree, 18 for Master's degree, and 23 for a Doctoral degree. The result of univariate ANOVA revealed that the years of education did not differ ($p > 0.05$) across the four groups (the mean years of education for the YNH, ENH, EHI, and YNM groups were 17.3, 16.1, 16.4, and 16.4, respectively).

A spectrally shaped masking noise was created to simulate the average of the EHI listeners' quiet thresholds and this background noise was introduced only to the YNM listeners. The masking noise was produced and shaped by a one-third octave band graphic equalizer within Adobe Audition (Adobe Systems Incorporated, San Jose, CA, version 1.5), using critical ratio estimation (Humes et al., 1987). Specifically, each mean pure-tone threshold of the EHI group was converted from dB HL to dB sound pressure level (SPL) in a 2-cm^3 coupler using ANSI (2004). Next, the critical ratio at each frequency was subtracted to estimate the noise spectrum level in dB SPL/Hz at each frequency that would produce a masked threshold equivalent to the mean quiet thresholds of the EHI group. The spectrum of the masking noise was then adjusted

J. H. Lee and L. E. Humes: Fundamental frequency and onset differences

via Adobe Audition to produce a noise with the desired spectral shape. In the pilot testing, masked pure-tone thresholds from three additional young normal hearers were measured to verify a close match (within ± 3 dB at octave intervals from 125 to 8000 Hz) between the individual masked thresholds of the three normal hearers and the mean quiet thresholds of the EHI group. This goal was achieved for the three listeners when the intensity of masking noise was increased by 5 dB from the initial estimate and this adjusted level was used for all the YNM listeners during the testing.

All the participants were native English speakers, recruited from Indiana University and the local community in Bloomington, Indiana. All were paid for their participation in this study.

B. Stimuli

The speech materials used in this study were sentences from the CRM corpus (Bolia *et al.*, 2000). Each CRM sentence has the constrained form "Ready (call sign), go to (color) (number) now." In the corpus, 8 different talkers (talker 0–3 for four males, and talker 4–7 for four females) produced 256 CRM sentences resulting from the combination of 8 call signs (Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, Tiger), 4 colors (Blue, Green, Red, White), and 8 numbers (1–8).

To select the male target voice among the available talkers of CRM corpus, the averaged $F0$ value of each talker speaking the same six CRM sentences was determined by COLEA MATLAB code (Loizou, 2000) using an autocorrelation approach. The male talker #0 was found to have a monotonous characteristic in voice pitch (Mean $F0 = 99$, SD = 1.48) among talkers and talker #0 was eliminated. Because male talker #3 was reported to yield higher performance compared to other talkers (Brungart, 2001), talker #3 was eliminated. This left male talkers #1 and #2, with male talker #1 ($T1$) chosen as the target voice in our study. The mean $F0$ values of all 256 sentences (8 call signs × 4 colors × 8 numbers) spoken by $T1$ were then measured. Results from COLEA analysis showed that the mean $F0$ value was 115.8 Hz ranging from 101.7 to 147.9 Hz (SD = 6.4 Hz). This value was comparable to the mean $F0$ value of $T1$ of 118.3 Hz estimated for these same materials and talker by Allen *et al.* (2008) using Kay Elemetrics Computerized Speech Lab version 4500.

Listeners were monaurally presented pairs of CRM sentences spoken by $T1$, differing in $F0$ and temporal onset. First, the $\Delta F0$ between sentences was manipulated using a high-quality speech analysis-synthesis system, STRAIGHT (Kawahara *et al.*, 1999). The STRAIGHT program estimated the $F0$ contour of each of the 256 sentences spoken by $T1$ in a 1-ms frame and then resynthesized the $F0$ contour to be shifted corresponding to the target amount of $\Delta F0$. The STRAIGHT MATLAB code has been used to modify the average $F0$ of sentences in recent studies because STRAIGHT is known to successfully shift $F0$ while preserving the natural pattern of the $F0$ contour in the target sentence without evoking much change in the formants (Carroll and Zeng, 2007; Stickney *et al.*, 2007). For conditions with a $\Delta F0$ of 0

ST, the $F0$ contours of both target and competing sentences were not processed. To provide $\Delta F0$ of 3 or 6 ST between two sentences, the $F0$ of one sentence was scaled upwards along the ST scale (3 or 6 ST) with the natural $F0$ variation unchanged. This up-shift of $F0$ was applied half to the target sentences and the other half to the masker sentences. Given the mean $F0$ value of 115.8 Hz (range: 101.7–147.9 Hz) for the 256 unprocessed sentences, the average $F0$ was shifted to 137.7 and 163.8 Hz after introducing $F0$ shifts of 3 and 6 ST, respectively.

Second, when the pair of sentences differed in onset asynchrony, the target sentence arrived first and then the masker sentence was presented after an onset asynchrony of 0, 50, 150, 300, or 600 ms. In other words, the target sentence always preceded the competing sentence by the amount of temporal onset asynchrony, resulting in some initial portion of the target sentence being presented in isolation. Because each CRM sentence had the same structure which resulted in similar overall sentence durations, the onset asynchrony naturally led to a difference in stimulus offset. Especially, the maximum value of 600-ms asynchrony among the five onset asynchronies enabled the listeners to hear "Ready Baron" from the target sentence in isolation before the competing sentence began and to hear "(number) now" within the competing sentence after the end of the target-sentence presentation.

C. Spectral shaping

All the unprocessed and $F0$ raised sentences were spectrally shaped in this study. The purpose of the spectral shaping was to provide speech audibility of at least a 10 dB sensation level (SL) from 200 to 4000 Hz for all EHI individuals. In order to achieve this goal without peak clipping, the overall level of the CRM sentences was reduced by 7 dB and then spectrally shaped using a one-third octave band graphic equalizer within Adobe Audition. Figure 1 shows the relative differences in amplitude spectra of the concatenated wave files (without the 7-dB overall amplitude reduction) from all the CRM sentences before and after the spectral shaping. After the shaping was applied, all the sentences were equated to have the same root-mean-squared (rms) amplitude (within ± 1 dB). The experimenter verified the absence of peak clipping for all sentences using Adobe Audition, as well as the absence of distorted sound quality by informal listening. The identical spectral shaping was applied to all stimuli in this study, rather than individually tailoring the spectral shaping to the hearing thresholds of the listeners.

To verify that the rms long-term spectrum of the speech (and calibration noise) was at least 10 dB above every listener's thresholds from 200 to 4000 Hz, especially for the 15 EHI individuals, the pure-tone hearing thresholds of each EHI listener in dB HL were converted to dB SPL based using the correction factors from ANSI (2004) and the thresholds in dB SPL at octave intervals from 250 to 8000 Hz were interpolated to values at one-third-octave intervals. The speech audibility was then calculated by subtracting individual pure-tone hearing thresholds in dB SPL (re: 2-cm^3 coupler at one-third-octave bands) from the rms
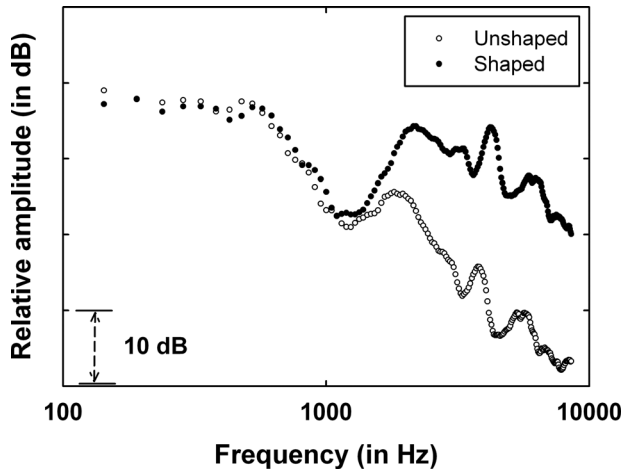
FIG. 1. Relative amplitude spectra of concatenated CRM wave files with or without spectral shaping.

spectrum of calibration noise at one-third-octave bands from 100 to 8000 Hz. Five EHI listeners who had the greatest hearing loss at 4000 Hz (EHI #1, 2, 3, 5, 15 shown in Table I) needed a presentation level greater than 85 dB SPL in order to provide speech audibility greater than 10 dB above the thresholds at 4000 Hz. Thus, those 5 EHI participants were presented speech materials at a presentation level of 91 dB SPL.

Figure 2 depicts the average SL of the speech stimulus at one-third-octave bands from 200 to 4000 Hz for the ENH group (unfilled circles), as well as 10 EHI listeners (unfilled triangles) tested at 85 dB SPL, and the 5 EHI listeners (filled triangles) tested at 91 dB SPL. For comparison purposes, two horizontal lines have been added to Fig. 2, one at 10 dB (dotted) and one at 15 dB (dashed) SL. A 10-dB SL through 4000 Hz was the minimum amount of speech audibility targeted for all the listeners, even with the EHI individual with the greatest hearing loss in this study. A 15-dB SL shown by the dotted line is optimal according to the Speech Intelligi-
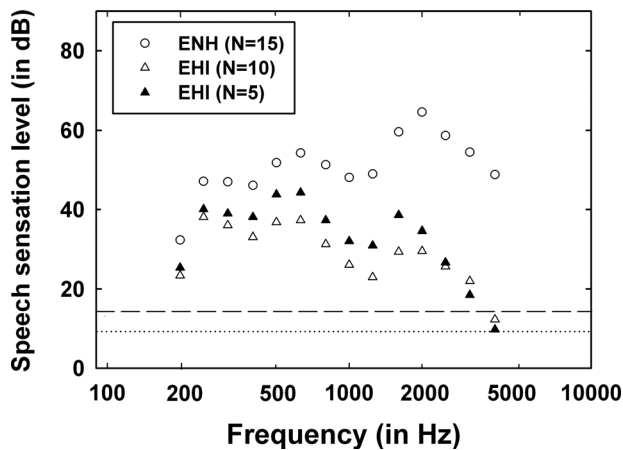

FIG. 2. Sensation level of the CRM speech signals presented to the 15 ENH listeners (unfilled circles), the 10 EHI listeners (unfilled triangles) tested at 85 dB SPL, and the 5 EHI listeners (filled triangles) tested at 91 dB SPL. The dashed line shows 15-dB SL displaying the optimal or asymptotic band SL according to the SII (ANSI, 1997) and the dotted line shows the objective of at least 10 dB SL for this study.

bility Index (SII) (ANSI, 1997). The EHI individual with the greatest hearing loss in the present study would hear the target speech with a sensational level greater than 16 dB from 100 to 3000 Hz (31-dB sensational level at maximum) and a sensational level of 10 dB at 4000 Hz.

D. Calibration and apparatus

For the purposes of calibration, a steady-state speech-shaped noise was created using Adobe Audition. This calibration noise was shaped to match, in terms of both long-term average spectrum and the average rms (within ± 3 dB), a concatenated wave file consisting of all the spectrally shaped CRM sentences spoken by talker $T1$. The match of long-term average spectrum and the average rms amplitude values was determined by an averaged (50-ms window) fast Fourier transform (FFT) analysis (Hanning window, FFT size = 1024).

The calibration noise was played through one channel of a 16-bit digital-to-analog converter (TDT DA1) at a sampling rate of 48 828 Hz and routed through an anti-aliasing filter (TDT FT5) with a cut-off frequency set to 10 kHz. The amplitude of the noise was set to 18 dB below maximum using a headphone buffer (TDT HB7) and sent to an insert phone (ER-3 A) coupled to an HA-2 2-cm^3 coupler (ANSI, 2004). Overall SPL and one-third- octave band levels were then measured in an HA-2 2-cm^3 coupler with a sound level meter (Larson-Davis, Provo, UT, model 800B) using a one-third octave band filter and a linear setting. With 18-dB attenuation at the headphone buffer, the overall presentation level of the noise was 85 dB SPL.

E. Procedures

Each participant was seated in front of a 17-in. touch-screen computer monitor in a single-walled sound-attenuating booth (Industrial Acoustics Company, Bronx, NY, Model 1200 A). The ambient noise levels in this booth complied with ANSI (1999) guidelines for threshold testing with earphones. All the participants listened to pairs of CRM sentences, one target and one masking sentence, with 0-dB TMR based on the average rms amplitude of the sentences. A custom MATLAB program was designed to add two competing CRM sentences digitally and to deliver them through the TDT System-III equipment (one channel of a 16-bit digital-to-analog converter). The customized MATLAB code also mixed the background noise required to simulate hearing loss in the YNM listeners with the sentence pairs. The simulation masking noise always preceded the onset of the first sentence by 150 ms and followed the offset of the second sentence in each pair by 150 ms.

After administering the screening tests for hearing and cognitive function, the participants were given written and oral instructions indicating that there would be two competing sentences presented, one of which contained the call sign of Baron, which identified the target sentence. After the presentation of sentence pairs in each trial, the listeners were required to make two responses. First, they needed to select YES/NO for whether they detected the presence of the cue word (call sign Baron). Second, regardless of their first

responses, they identified the color and number in the sentence containing the cue word. A touch screen allowed listeners to make both detection and identification responses as well as allow the subjects control time to choose their responses. To examine the detection sensitivity to the cue word Baron for the first task, the cue word Baron was contained in 75% of the trials ($N = 24$) in each block. Participants were informed that Baron would be presented on only 75% of the trials, but were not informed that the target sentence always arrived earlier than the masking sentence for the onset asynchrony.

Prior to the experimental testing, practice and familiarization sessions were given. In the practice session, the listeners heard 32 target sentences without a competing sentence and they needed to select YES/NO for the presence of Baron as well as the CN responses. All the listeners could detect Baron with 100% accuracy and could also identify the CN sequence with greater than 98% accuracy, verifying that all the keywords in quiet were sufficiently audible and identifiable to all listeners. Following the practice session, the subjects were given a long familiarization session in which the same (detection and identification) tasks, the 75% proportion of Baron trials ($N = 90$) and 25% of "No-Baron" trials ($N = 30$), were preserved. Only during the familiarization task trial-to-trial feedback was visually shown immediately after the listeners' responses.

In the experimental testing, a total of 1920 trials were presented to each listener (32 trials × 4 blocks × 15 conditions) as the 15 listening conditions were generated by a combination of three values of $\Delta F0$ (0, 3, 6 ST) and 5 levels of onset asynchrony (0, 50, 150, 300, 600 ms) between the two sentences. The test sequence of the 15 conditions for each listener was a quasi-randomized order. Each experimental session lasted approximately 1.5–2 h. To complete all testing, three to four sessions and four to five sessions were required for young and elderly listeners, respectively. While the speech signals were presented to the test ear only through the insert earphone, the non-test ear was also occluded by the other insert earphone during testing.

F. Scoring and data analysis

Recall that the tasks of the listeners were to select the "YES" or "NO" response regarding whether they detected the cue word Baron and then to choose the color and number spoken by the target voice identified by Baron. From the detection responses, the sensitivity, d' to detect the cue word was estimated based on the signal detection theory. From the identification answers, the number of correctly or incorrectly identified color and number responses was separately scored. Further details regarding the scoring are as follows.

1. Scoring for the detection task

Performance for detection sensitivity, d', was determined based on the proportion of "Hit" and "False alarm" responses summed across the four blocks (128 trials) for each condition. The proportion of Hit and False alarm responses was converted into z scores, where the proportion

of Hit responses was the summed number of Hit responses divided by the number of Baron trials ($N = 96$) and the proportion of False alarm responses means the aggregated False alarm responses relative to the total "No-Baron" trials ($N = 32$). The difference between the z-transformed probabilities of Hit and False alarm is d' (Green and Swets, 1966; Macmillan and Creelman, 2005). The Hit and False alarm rates of 0 or 1 were adjusted to be 0.004 or 0.996 (Stanislaw and Todorov, 1999) resulting in a maximal d' of 5.3, meaning that the detection performance was perfect (100% Hits, 0% False alarms).

2. Scoring for the identification task

The CN identification was scored correct only when both the color and number responses were correct. Because the listeners were instructed to identify the color and number spoken by the target voice saying Baron, their identification accuracy was scored only when the presence of Baron was correctly detected (i.e., on trials yielding Hit). In other words, we did not evaluate the CN responses when the detection response of Baron was a "Miss," False alarm, or "Correct rejection."

Given that the $F0$ shift was applied half of the time to the target sentences and half of the time to the masker sentences, additional paired t-tests were conducted to examine whether it mattered whether the shift was applied to the target or masker sentence. Given no significant difference, the CN identification responses were averaged across the two cases (target-shift and masker-shift), representing the mean CN score of each of the 15 conditions. An analysis of partially correct and completely incorrect responses was also performed to examine "intrusions" of the competing CN stimulus.

3. Data analysis

The current research design with a total of 15 conditions included 1 between-subject (YNH, ENH, EHI, and YNM) and 2 within-subject factors ($F0$: 0, 3, 6 STs; Onset asynchrony: 0, 50, 150, 300, and 600 ms). All the individual proportion-correct scores for CN identification were converted into rationalized arcsine units (RAU) to stabilize the error variance (Studebaker, 1985) prior to statistical analyses. A 4 × 3 × 5 factorial mixed-model ANOVA, with 4 between-group and 2 within-group comparisons, was conducted on each of the 3 dependent measures: d', correct CN identification score, and intrusions.

The mean d' values were compared to examine the effects of group, $\Delta F0$, and onset asynchrony on the detection sensitivity. The data for CN identification accuracy, as well as the proportion of intrusions among incorrect CN responses, were also analyzed for the same main effects. Interactions among group $\Delta F0$ and onset asynchrony were also examined for each of these four dependent values. Any necessary *post hoc* multiple comparisons were conducted based on the adjusted criterion p value, depending on the number of paired comparisons to be made. For analysis, the Greenhouse–Geisser correction was used when the Mauchly's Test of Sphericity was violated.

III. EXPERIMENT 1: RESULTS AND DISCUSSION

A. Detection sensitivity (d')

For the detection task, all the listeners were required to respond Yes or No for the detection sensitivity (d') of the cue word Baron. Figure 3 shows mean and standard errors of d' for Baron as a function of onset asynchrony for the four groups when two sentences were separated by three levels of $\Delta F0$: 0 (left), 3 (center), and 6 (right) STs. As previously noted, the maximum possible d' was 5.3 (where hit rate of 1 and false alarm rate of 0 were adjusted to be 0.996 and 0.004) and the minimum possible d' was 0 (chance level). Without $F0$ and onset separation (0-ST $\Delta F0$ and 0-ms onset asynchrony), mean d' was 2.65, 2.27, 2.3, and 2.27 for YNH, ENH, EHI, and YNM listeners, respectively, indicating that Baron was similarly detected for all listeners. Averaged d' values collapsed across Δonsets and groups were 3.51, 3.75, 3.96 for $\Delta F0$s of 0, 3, 6 ST, respectively. Mean d' values collapsed across 3 $\Delta F0$s and 4 groups progressively increased from 2.58, 3.22, 3.49, 4.52 to 4.89 when the onset asynchrony increased from 0, 50, 150, 300 to 600 ms, respectively.

A $4 \times 3 \times 5$ factorial mixed-model ANOVA was performed on d' values with a between- subjects factor of group and two repeated-measures variables ($\Delta F0$ and Δonset). Results showed that detection sensitivity of Baron was significantly ($p < 0.01$) affected by $\Delta F0$ [$F(2, 112) = 30.4$] and onset asynchrony [$F(3.2, 179.1) = 187.1$], but not by group [$F(3, 56) = 0.4$]. This demonstrates that the overall detection sensitivity of Baron was comparable across the listener groups and that all the listener groups detected Baron more readily as the differences in $F0$ and onset between two sentences increased. The results from Bonferroni-adjusted multiple paired-comparisons showed significant improvements in d' of about 0.2–0.3 for each successive increment in $\Delta F0$ values. Also, the detection performance at the 0, 50, 150, 300, and 600-ms onset asynchronies significantly differed from one another. In particular, the d' value to detect Baron greatly increased between 150 and 300 ms asynchrony at 0-ST $F0$ difference, as visually seen in Fig. 6, although the cue word Baron still temporally overlapped with the competing message up to 300-ms asynchrony. Except for the two-
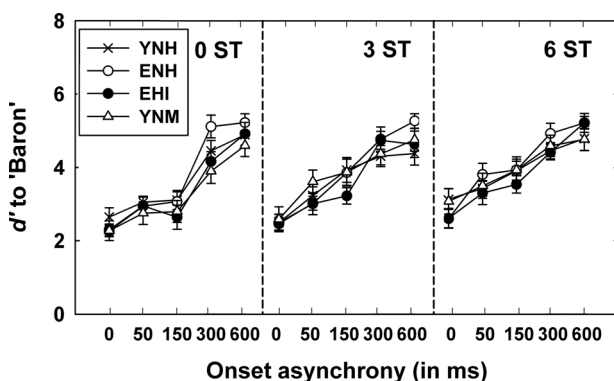
way interaction between $\Delta F0$ and onset asynchrony, other interactions were not significant. The significant two-way interaction between $\Delta F0$ and onset asynchrony [$F(8, 448) = 3.6$, $p < 0.01$] is because of slightly lower d' values between 50 and 150 ms onset asynchronies at $\Delta F0$ values of 0 ST compared to those at 3 or 6 ST.

$F0$ and sentence onset separation cues significantly improved call-sign detection sensitivity in this experiment. Both $\Delta F0$ and Δonset segregation cues by themselves have been considered low-level physical segregation cues encoded in the periphery for perceptual organization in a multi-talker environment (Darwin, 2008). However, the processing of those primitive grouping cues represents processing by additional peripheral, central, and cognitive mechanisms, as suggested by the involvement of a multi-stage processing model (Alain, 2007; Alain et al., 2005; Snyder and Alain, 2005, 2007). Especially, Alain et al. (2001) showed that age-related deficits in detecting a mistuned harmonic depended not only on the listeners' peripheral factors but also upon an age-related decline in central auditory functioning. As noted, however, on a group basis there were no significant differences in call-sign detection among the four groups. This suggests that, if there were age-related central or cognitive deficits in the older listeners in this experiment, then they were insufficient to have an impact on call-sign detection.

Additional insights into this can be gleaned from examination of individual differences in call-sign detection and the association of these differences with age, hearing loss, and cognitive function. Correlations were examined between d' values averaged across 15 conditions and 3 individual factors: high-frequency (1, 2, 4 kHz) pure-tone thresholds, age, and digit span. No significant correlations were observed. Given this non-significant relation and the comparable d' across four groups, aging or age-related declines in the short-term working memory evaluated by digit span did not limit the detection performance of the older subjects, at least for the closed-set CRM task in this experiment.

In summary, we found that the $\Delta F0$ and onset asynchrony significantly improved the sensitivity (d') measures. Even when the listeners heard two CRM messages without $F0$ and onset asynchrony difference, detection performance for Baron was good (the average d' was >2.0 throughout) and equivalent across the four groups.

B. Correct CN identification

In the current study, after responding Yes or No for the detection of Baron, listeners needed to choose both the color and the number spoken by the talker who produced the cue word Baron. CN identification scores in percent-correct were converted to RAU, and the means and standard errors of the CN identification accuracy (in RAU) in each of the 15 conditions (3 $\Delta F0$s $\times$ 5 onset asynchronies) are displayed in Fig. 4 for the 4 listener groups. Identification scores are plotted as a function of onset asynchrony for $\Delta F0$ values of 0 (left), 3 (center), 6 (right) STs. When the listeners heard two CRM messages without $F0$ and onset separation ($\Delta F0 = 0$ ST, Δonset $= 0$ ms), the mean CN identification performance of

FIG. 3. Detection sensitivity (d') to cue word Baron as a function of $\Delta F0$ and onset asynchrony for the four listener groups (YNH, ENH, EHI, and YNM listeners).
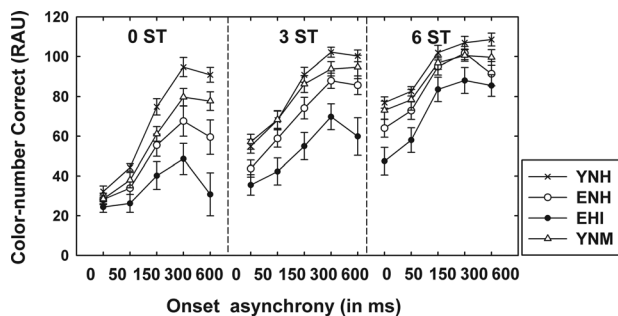
FIG. 4. CN identification scores in RAUs as a function of $\Delta F0$ and onset asynchrony for the four listener groups (YNH, ENH, EHI, and YNM listeners).

the four groups was similarly poor (mean CN accuracy of 32, 28, 24, and 29 RAU for the YNH, ENH, EHI, and YNM listeners, respectively). The relatively poor CN scores of the listeners were comparable to earlier reports for young and older adults that the mean CN identification accuracy ranged from 25% to 35% when the competing CRM sentences spoken by the same talker or same-gender talkers were monaurally and simultaneously presented (Humes and Coughlin, 2009; Humes et al., 2006; Rossi–Katz and Arehart, 2009).

CN identification values in RAU were subjected to a $4 \times 3 \times 5$ ANOVA, with group as a between-subjects factor and $\Delta F0$ and onset asynchrony as two repeated-measures factors. The statistical results showed that the main effects of group, $\Delta F0$, and onset asynchrony were all significant ($p < 0.01$). This demonstrates that both the $\Delta F0$ and onset separation between two competing voices substantially enhanced the identification of the target color and number, and that CN identification was different across listener groups. The between-subjects variable group significantly ($p < 0.01$) interacted with the repeated-measures variables $\Delta F0$ [$F(5.4, 100.8) = 4.8$] and Δonset [$F(8.5, 158.2) = 4.5$]. The $\Delta F0 \times \Delta$onset interaction [$F(5.3, 299.2) = 4.0$] and the three-way interaction of group $\times \Delta F0 \times \Delta$onset [$F(16, 299.2) = 5.0$] were also significant. As a result, several *post hoc* analyses were conducted in order to analyze those significant main effects and interactions more closely.

First, 3 ($\Delta F0$) × 5 (onset asynchrony) repeated-measures ANOVAs and, as needed, Bonferroni-adjusted multiple paired-comparisons, were conducted separately on the data from each listener group. Results showed that the main effect of $\Delta F0$ was significant within each group, and the CN identification improved with increasing $\Delta F0$. When onset asynchrony was not provided between sentence pairs, the degree of $\Delta F0$ benefit resulting from a 3 ST shift in $F0$ (performance score for the 3 ST, 0 ms condition minus score for the 0 ST, 0 ms condition) was, on average, 23, 16, 11, and 29 percentage points for the YNH, ENH, EHI, and YNM groups, respectively. Similar to the present study, Darwin et al. (2003) presented competing CRM sentences to young normal hearers at a 0-dB target-to-competition ratio, and found that listeners' CN identification scores improved by about 20 percentage points for a $\Delta F0$ of 4 ST. Brokx and Nooteboom (1982) also reported a comparable improvement of 20% (from 40% to 60% correct) in content word identification performance when a $\Delta F0$ value of 3 ST separated

monotonous target sentences from continuous background speech. In our data, the mean $\Delta F0$ benefit from separation of 6 ST was 45, 37, 23, and 44 percentage points for the YNH, ENH, EHI, and YNM groups, respectively. Thus, the mean $\Delta F0$ benefit for both 3 and 6 ST in the EHI group was almost half of the mean benefit in the young groups, even when one young group (YNM) had comparable inaudibility of high frequencies.

Second, the main effect of onset asynchrony was also significant within each group, and CN identification was enhanced with a greater onset separation of the two competing messages (performance for 300 ms > 150 ms > 50 ms > 0 ms, all significant at $p < 0.01$). When $\Delta F0$ was 0 Hz, the benefit from a 50-ms onset asynchrony for the identification performance (the score for the 0 ST, 50 ms condition minus the score for the 0 ST, 0 ms condition) was, on average, 12, 6, 2, and 10 percentage points for the YNH, ENH, EHI, and YNM groups. Thus, the brief period of 50-ms onset-asynchrony, which was the minimum onset separation manipulated in the current study, improved the CN identification accuracy for young adults more than for the elderly. The mean benefit from the 600-ms onset asynchrony was 55, 32, 11, and 48 percentage points for YNH, ENH, EHI, and YNM groups, respectively. The EHI group consistently received less benefit than the other groups from onset asynchrony, independent of the amount of onset asynchrony. This reduced onset benefit of listeners who have cochlear pathology was somewhat consistent with the finding of HI listeners (both young and older adults) in Lentz and Marsh (2006). Lentz and Marsh (2006) also found a reduced benefit for onset asynchronies from 100 to 300 ms in their HI listener group ($N = 7$, age = 25–61 yrs) than in their NH group ($N = 7$, age = 18–51 yrs). Given no significant relation between the amount of onset-asynchrony benefit and the audible frequency range in the excitation-pattern of their HI listeners, the use of onset asynchrony was assumed to be associated with a suprathreshold or more high-level processing, such as temporal integration or vowel-dominance perception (Lentz and Marsh, 2006). The results from the present study also suggest that audibility is not the sole underlying factor limiting the performance of EHI subjects. Rather, the combination of cochlear pathology and aging appears to be responsible. If audibility alone were key, then there would have been few performance differences between the YNM and EHI groups. In addition, the stimuli in this study were spectrally shaped to minimize audibility deficits.

Although all listener groups benefited from $\Delta F0$ and onset asynchrony, as can be seen in Fig. 4, the improvement with increasing cue size was smaller for both groups of older adults, but especially for the EHI group. Note that all four groups showed similar CN identification performance of 24–32 RAU in the hardest condition (0 ST, 0 ms), yet a different interaction pattern was observed across groups. That is, the onset benefit was fairly uniform across $\Delta F0$ values for the older groups whereas the young groups had less onset benefit at $\Delta F0$ of 6 ST than at $\Delta F0$ of 0 ST, indicating that the young groups' identification performance depended more on the availability of other cues. Especially, the EHI group experienced the least amount of improvement in

performance with increasing onset asynchrony alone or increasing $\Delta F0$ alone. Overall, the smaller benefit observed for the elderly agrees with the findings of Humes and Coughlin (2009) in which, using the same speech material (CRM corpus), older adults benefited less from improved listening conditions, including greater acoustical differences between competing talkers, despite equivalent performance of young and older adults in the more difficult baseline condition.

Next, several *post hoc* univariate ANOVAs and *t*-tests were conducted using CN identification scores in RAU in each of the 15 conditions in order to examine the main effect of the group for each of the 15 conditions. If significant effects of the group were observed in ANOVA, *post hoc* *t*-tests were performed to identify the specific group differences. Given the 15 comparisons, the adjusted *p* value of 0.00067 (0.01/15) was used as the criterion for significance. Statistical results showed that none of the conditions revealed significant group differences between the ENH and EHI, between the ENH and YNH, and between the YNH and YNM groups, revealing no effect of high-frequency hearing sensitivity alone or of aging alone on the identification performance. In fact, all significant differences between groups involve the EHI group performing significantly worse than either or both of the two young groups. Thus, overall, the combined effects of cochlear pathology and aging appear to lead to the poor performance of the EHI listeners on these competing-speech tasks. This result is supported by Snyder and Alain (2007) who reported that segregation of sounds is likely to occur beginning in the auditory periphery and continuing to the primary or secondary auditory cortex, depending on the complexity of cues.

To examine individual differences in CN identification among the older adults, the data for the ENH and EHI subjects were pooled, and correlations between mean CN scores across 15 conditions and 3 subject factors (high-frequency thresholds, age, and digit spans) were computed. The averaged CN accuracy of elderly individuals was significantly ($p < 0.05$), negatively, and moderately correlated with high-frequency thresholds ($r = -0.50$), and also significantly, positively, but somewhat weakly, correlated with digit spans ($r = 0.37$). Like the current study, Humes *et al.* (2006) also measured CRM-task performance of EHI listeners with or without manipulating uncertainty, when selective- or divided-attention was required. Among four predictor variables observed [average digit-span score, age, hearing loss asymmetry, and average high-frequency hearing loss (1, 2, 4 kHz)], only the digit span score was significantly correlated with CRM performance of EHI individuals, regardless of, or the degree of, uncertainty or the attention type. In Humes *et al.* (2006), however, the correlations with digit span were greatest for the divided-attention conditions, which required subjects to hold both sentences in memory prior to being prompted with the cue for the target sentence. As a result, the correlations between performance and digit span in that study were somewhat higher than observed in this study. Nonetheless, the correlations observed here support the combined influence of cochlear pathology and cognitive function on the identification performance of elderly individuals, consistent with previous observations (e.g., Humes and Coughlin, 2009; Humes *et al.*, 2006; Rossi–Katz and Arehart, 2009; Snyder and Alain, 2007; Tun *et al.*, 2002).

Finally, we also evaluated the relation between mean d' and mean CN values averaged across the 15 conditions for the young listeners ($N = 30$) and the elderly listeners ($N = 30$), separately. Recall that all the listeners were required to respond Yes or No for the detection task of Baron. Because this cue word is considered a lexical cue identifying a target message between two competing messages, the inaudibility of Baron could evoke difficulty in segregating the two voices or messages which would then lead to difficulty in identifying the target message spoken by the target voice (Shafiro and Gygi, 2007). The results of correlation analyses revealed significant ($p < 0.01$) correlations between d' and CN identification performances for both young and older groups (Young: $r = 0.58$; Elderly: $r = 0.63$). This supports the idea of Shafiro and Gygi (2007) that the inability to detect Baron could lead to difficulty in segregating the two voices or messages, leading to difficulty identifying the target message spoken by the target voice, regardless of listeners' age.

In summary, our data CN identification revealed that: (1) Identification of the color and the number words for each of the four listener groups was substantially enhanced from $F0$ and onset differences between two messages; (2) for the EHI group, the benefit of $\Delta F0$ and onset-asynchrony cues was smaller compared to other groups; (3) combined effects of cochlear pathology and age-related cognitive function seem to contribute to the poorer identification performance of EHI individuals; and (4) independent of age, all the listeners who better detected the cue word Baron, which was the cue identifying the target voice, actually better identified the target message against the competing one.

C. CN intrusions among incorrect responses

Other than correct CN identification responses, we also explored how often the incorrect responses occurred from the competing CN information. The number of intrusions in each of the 15 two-talker conditions across the four groups was counted, where an intrusion was defined as incorrect identification responses corresponding to the competing voice's color, number, or both. This analysis of misidentified responses examined whether the source of incorrect responses was confusion with the competing message information (informational masking) or a random guessing response due to inaudibility of both messages (energetic masking).

Figure 5 plots the proportion of intrusions in RAU relative to the total incorrect responses as a function of onset asynchrony across the four groups. Recall that one out of the four colors and one of the eight numbers were chosen for the closed-set identification response format. Because the competing CRM sentences did not have the same color or number, the three colors and seven numbers remaining could be selected with equal likelihood. Approximately 52% of incorrect responses (51.9 in RAU) could be randomly selected from the competing coordinates [i.e., a summation of 1/3 chance (33%) from the 3 remaining non-target colors, 1/7
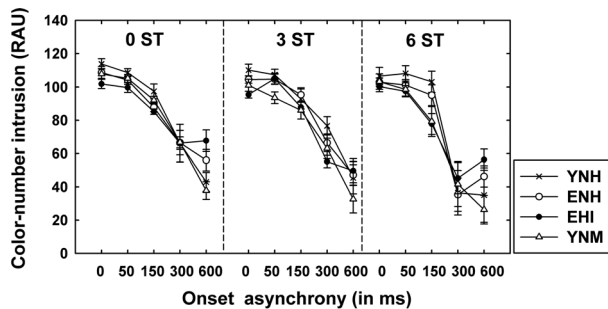
FIG. 5. CN intrusions in RAUs as a function of $\Delta F0$ and onset asynchrony for the four listener groups.

chance (14%) from the 7 remaining non-target numbers, and 1/21 chance (5%) for guessing both the color-number coordinates of the competing sentence].

As illustrated in Fig. 5, the intrusions dominated most (>90%) of the incorrect responses when two CRM sentences competed without $F0$ and onset differences (0 ST, 0 ms condition), supporting that a greater similarity between competing signals yields a high degree of informational masking (Brungart, 2001; Brungart et al., 2001; Darwin et al., 2003; Humes et al., 2006; Rossi–Katz and Arehart, 2009; Srinivasan and Wang, 2008). Further, for onset asynchronies of 0, 50, and 150 ms, as well as $\Delta F0$ alone, the vast majority of errors ($\geq$80%) are still intrusions.

A $4 \times 3 \times 5$ multivariate ANOVA was performed to investigate the effects of group, $\Delta F0$, and onset asynchrony on the proportion of intrusions. The proportion of intrusions was significantly reduced as $\Delta F0$ and onset asynchrony increased [$\Delta F0$: $F(1.7, 95.7) = 14.7$; Δonset: $F(2.3, 129.4) = 193$]. A two-way interaction ($\Delta F0 \times \Delta$onset) was also significant [$F(4.3, 243.2) = 4.8$], such that the reduction of intrusions with onset asynchrony was greater for larger $\Delta F0$ values. This confirms that a larger $F0$ and onset separation facilitates segregation of audible competing voices, consequently decreasing the confusion with the competing CN coordinates.

Interestingly, results revealed that the intrusion rate did not significantly differ across the four groups [$F(3, 56) = 2.6$], even though the older group tended to make more errors. That is, the older adults identified competing CN words less accurately than the young adults, but the source of incorrect responses seemed to be similar across the groups. Especially at 600-ms onset separation, the older listeners tended to have a relatively greater proportion of intrusions (53.9 RAU) compared to the young groups (36.7 RAU). Recall that at the 600-ms onset asynchrony, the (number) now portion of the competing sentence no longer overlapped temporally with the target sentence. The number portion of the competing CN coordinates was therefore presented last and in the clear. The 600-ms delay was presumed to be more beneficial than 300-ms of onset separation due to a less temporal overlap, but this was not the case for the EHI group. Although EHI listeners showed a slight decrease in CN identification performance from 300 to 600 ms, the proportion of intrusions increased as onset asynchrony increased from 300 to 600 ms even with $\Delta F0$ of 6 ST. A possible reason for this is that the elderly listeners may have a limited

ability to suppress or inhibit the competing message compared to the young groups despite a match of speech audibility between two groups. The finding supports an age-related inefficiency of an inhibitory mechanism (Hasher et al., 1999, 2007; Hasher and Zacks, 1988; Sommers and Danielson, 1999; Wright and Elias, 1979), yielding reduced selective attention to the target message and ineffective suppression of the distracting message in elderly listeners.

Taken together, all the listeners had difficulty identifying the target-message content when a small difference of $F0$ and onset separated the two competing sentences. A greater dissimilarity between the two sentences not only substantially increased identification performance but also significantly reduced the confusion with the competing information. As long as the audibility of competing messages was restored, the source of incorrect responses appeared to be equivalent across the four listener groups.

IV. EXPERIMENT 2: RATIONALE AND METHODS

In Experiment 2, the trial-to-trial stimulus uncertainty was maximized considerably compared to Experiment 1. In Experiment 1, for a given condition involving a difference in $F0$, the $F0$ shift was applied to 1 talker, either the target or competing talker, throughout a given block of 32 trials. With regard to onset asynchrony, consistent with much of the prior research on this cue (Hedrick and Madix, 2009; Lentz and Marsh, 2006), the order of the two competing sentences in a pair was fixed such that the target sentence was always presented first. Thus, in addition to the cue word Baron marking the target sentence, the first occurring sentence was always the target sentence. Thus, there were ultimately two cues to the target sentence in low-uncertainty Experiment 1: Baron and the first of two sentences.

As described earlier, Brungart and Simpson (2004) found that randomizing the masker had little effect on young listeners' CN identification whereas the trial-to-trial randomization significantly influenced older listeners' performance (Humes et al., 2006). Recently, Mackersie et al. (2011) found that trial-to-trial uncertainty in the target $F0$ significantly affected CN identification for NH and HI listeners, but not for all conditions.

Given the inconsistent impact of uncertainty across listener groups, as well as a few studies of the effects of uncertainty on the use of $\Delta F0$ or onset segregation cues, Experiment 2 examined the influence of uncertainty on performance. In particular, this experiment examined whether increasing uncertainty by randomly applying the onset asynchrony to either the target or masker sentence impacted performance relative to that observed in Experiment 1. We wanted to know the effects of uncertainty on d' and CN performance, but also on the relative influence of $F0$ and onset asynchrony for young and older adults (YNH, ENH, EHI).

It is hypothesized, based on the results of Experiment 1, that trial-to-trial uncertainty would decrease the detection of the cue word Baron, as well as overall CN identification performance. This would be most noticeable for the onset-asynchrony cue since the application of the onset-asynchrony cue was fixed throughout Experiment 1 (target always first),

whereas the application of the $\Delta F0$ cue alternated from block to block (but was fixed within a block).

A. Methods

1. Listeners

Twenty-four listeners (8 YNH, 8 ENH, and 8 EHI) who did not take part in Experiment 1 participated in Experiment 2. All the listeners were native English speakers, recruited from Indiana University and the local community in Bloomington, Indiana, and all were paid for their participation.

The ages of the YNH adults ranged from 19 to 26 yr ($M = 22.6$, SE = 0.84 yrs). The average age of the ENH adults was 70.5 yr (SE = 1.95 yrs, ranging from 64 to 82) whereas the average age of the EHI adults was 74.3 yr (SE = 2.15 yrs, ranging from 66 to 81). As in Experiment 1, all the listeners had normal middle-ear status, a score of 25 or greater on the MMSE for cognitive status, and a score of at least 9 on the summed result of the auditory forward and backward digit-span for memory. The mean self-reported education levels were 17, 15.25, and 15.75 yrs for the YNH, ENH, and EHI groups, respectively. Results of pair-wise independent sample t-test and one-way ANOVAs indicated that the ENH and EHI adults did not differ in age and that the three listener groups did not differ significantly ($p > 0.05$) in MMSE [$F(2, 21) = 0.46$], digit-span scores [$F(2, 21) = 0.68$], or education level [$F(2, 21) = 1.5$].

All 8 YNH listeners had air-conduction hearing thresholds better than 20 dB HL (ANSI, 2004) at octave frequencies from 250 through 8000 Hz. The mean thresholds of the ENH participants were better than 20 dB HL at octave frequencies from 250 through 4000 Hz and their averaged hearing threshold at 8000 Hz was 23.8 dB HL. EHI listeners' mean thresholds were 21, 24, 26, 31, 53, 66 dB HL at octave frequencies from 250 through 8000 Hz. Like Experiment 1, Experiment 2 also avoided overlap in the range of air-conduction thresholds between the ENH and EHI individuals at frequencies from 1000 to 8000 Hz. Test ear selection for the monaural presentation was conducted as described in Experiment 1. In Experiment 2, the right ear was selected as the test ear for all 8 YNH, 6 of 8 ENH, and 4 of 8 EHI listeners.

2. Stimuli

The primary goal of Experiment 2 was to examine the effect of the trial-to-trial stimulus variability on CN identification performance. As explained earlier, in Experiment 1, the three $F0$ cues (0, 3, 6 ST) and the five onset asynchrony cues (0, 50, 150, 300, and 600 ms) were combined, resulting in 15 conditions. In Experiment 2, only two values of $F0$ difference (0, 6 ST) and two values of onset asynchrony (0, 300 ms) were used. This was designed to reduce the number of possible stimulus combinations from Experiment 1 and also to avoid floor or ceiling effects observed in that experiment. Since the cues could now be applied to the target sentence only, the masker sentence only, or both target and masker sentences, nine conditions resulted as shown in Table II. For each of these 9 stimulus conditions, 128 trials were presented (32

trials × 4 blocks per condition) as in Experiment 1, but the complete set of 1152 trials was presented in a completely randomized fashion in Experiment 2. This large set of trials, however, was administered in blocks of 32 trials to allow for a sufficient number of breaks and to match that feature of stimulus presentation from Experiment 1.

3. Procedures and data analysis

As in Experiment 1, each sentence was presented to the test ear at an overall level of 85 dB SPL, and the non-test ear was similarly occluded. Testing devices and environment were also the same. The same tasks (call-sign detection and CN identification) were required for all the participants after the completion of the screening tests for hearing and cognitive function (MMSE and digit span tests) and the practice session. The result of the practice session presenting target sentences without competition showed that all the listeners could detect Baron correctly and could identify the color and number words with greater than 98% accuracy in quiet.

Listeners also completed a familiarization session to minimize learning effects. The familiarization session in Experiment 2 presented 72 trials, consisting of 8 trials representing each of the 9 conditions. As in Experiment 1, the listeners received correct-answer feedback only during the familiarization session. Also, as in Experiment 1, Baron occurred on 75% of the trials for both familiarization and experimental sessions. For the completion of all testing in Experiment 2, participants needed two to three sessions (1.5–2 h per session).

The methodology used for scoring was the same as for the previous experiment. For data analysis, data from the nine conditions (see Table II) of Experiment 2 were collapsed for each $\Delta F0$ or Δonset cue, regardless of whether the cues were applied to the target or masker sentences. This resulted in a total of four stimulus conditions for data analysis: (1) $\Delta F0 = 0$ ST, Δonset $= 0$ ms; (2) $\Delta F0 = 0$ ST, Δonset $= 300$ ms; (3) $\Delta F0 = 6$ ST, Δonset $= 0$ ms; and (4) $\Delta F0 = 6$ ST, Δonset $= 300$ ms.

B. Results and discussion

1. Detection sensitivity (d')

To investigate the effects of trial-to-trial variability on the use of $\Delta F0$ or onset separation cues in Experiment 2,

TABLE II. Nine conditions generated by two levels of $\Delta F0$ or Δonset alone, and also four possible combinations of each. The application of segregation cues to either target or masker is shown in parentheses.

	$\Delta F0$	Δonset
Condition 1	0 ST	0 ms
Condition 2	6 ST (target)	0 ms
Condition 3	6 ST (masker)	0 ms
Condition 4	0 ST	300 ms (target)
Condition 5	0 ST	300 ms (masker)
Condition 6	6 ST (target)	300 ms (target)
Condition 7	6 ST (masker)	300 ms (masker)
Condition 8	6 ST (target)	300 ms (masker)
Condition 9	6 ST (masker)	300 ms (target)

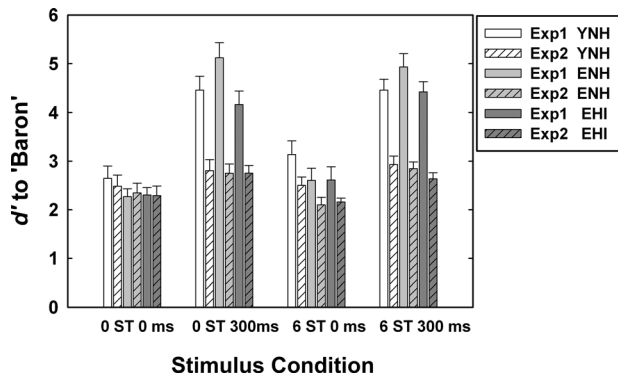J. H. Lee and L. E. Humes: Fundamental frequency and onset differences

FIG. 6. Detection sensitivity (d') of Baron obtained from YNH, ENH, and EHI listeners of Experiment 1 (solid) and Experiment 2 (striped) for the four stimulus conditions. Abscissa labels show the $\Delta F0$ value in STs and the onset asynchrony in milliseconds (ms) for each stimulus condition.
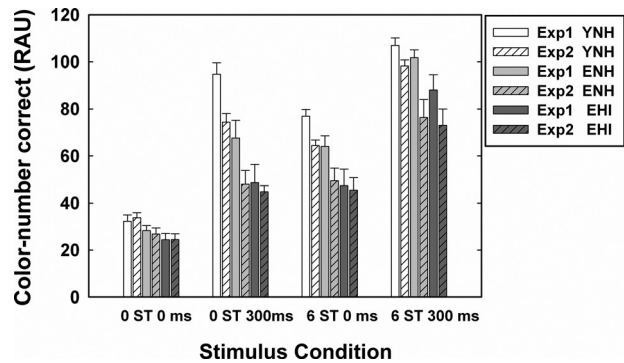

FIG. 7. CN identification (in RAU) of YNH, ENH, and EHI listeners in Experiment 1 (solid) and Experiment 2 (striped) for the four types of stimulus conditions. Abscissa labels show the $\Delta F0$ value in STs and the onset asynchrony in milliseconds (ms) for each stimulus condition.

Fig. 6 compares detection sensitivity, d', obtained from the two experiments. The striped bars display data for Experiment 2 and the solid bars for Experiment 1 for the same $\Delta F0$ and onset-asynchrony values and for the same three listener groups (YNH, ENH, and EHI).

When $\Delta F0$ and onset separation cues were not present (0 ST/0 ms), the d' values for all the listeners were similar (ranging from 2.2 to 2.7) across the two experiments. For the other three stimulus conditions, however, the cue word Baron was considerably more difficult to detect during complete randomization (Experiment 2) than when being fixed (Experiment 1). This is particularly true when onset asynchrony was included as a cue. In fact, in Experiment 2, the detectability of the cue word is now roughly the same across all four stimulus conditions with d' values ranging between about 2.2 and 3.0 across conditions and groups.

The d' values in Experiment 2 were subjected to a 3×4 mixed-model ANOVA with a between-subjects factor of group (YNH, ENH, and EHI) and within-subjects factor of stimulus condition (0 ST/0 ms, 0 ST/300 ms, 6 ST/0 ms, and 6 ST/300 ms). Results revealed that detection sensitivity of Baron was significantly ($p < 0.01$) affected by stimulus condition [$F(1.96, 41.1) = 19.2$], but not by group [$F(2, 21) = 0.62$]. The two-way interaction between group and stimulus condition was not significant. Results of Bonferroni-corrected multiple paired-comparison analyses showed that call-sign detection performance for the following paired comparisons were significantly different: d' for 0 ST/0 ms condition < 0 ST/300 ms or 6 ST/300 ms condition, d' for 6 ST/0 ms condition < 0 ST/300 ms or 6 ST/300 ms condition.

Additional correlational analyses were performed for the 16 elderly listeners to examine whether there were associations between d' values and high-frequency threshold, age, or digit spans. No significant correlations emerged among these 16 older listeners.

2. Correct CN identification

Figure 7 shows the mean and standard errors for the RAU-transformed CN identification performance from Experiment 1 (solid) and Experiment 2 (striped) for YNH, ENH, and EHI groups. In general, except for the most diffi-

cult baseline listening condition (0 ST/0 ms), performance was generally worse in Experiment 2 than in Experiment 1 across stimulus conditions and groups. The EHI listeners, however, tend to show the smallest differences in CN identification across experiments.

In order to examine the effect of group and stimulus condition on CN identification, a 3×4 ANOVA was performed and followed by *post hoc* Bonferroni-corrected multiple paired-comparison analyses. The results showed that CN identification was significantly ($p < 0.01$) affected by both group [$F(2, 21) = 9.8$] and stimulus condition [$F(1.95, 40.9) = 142.3$]. The two-way interaction between group and stimulus condition was not significant. Paired-comparison testing showed that CN identification performance for the 0 ST/0 ms condition was significantly worse than that in the other three conditions. Also, identification performance with onset separation of 300 ms alone or $\Delta F0$ of 6 ST alone was poorer compared to the performance for the stimulus condition presenting both of these cues. Interestingly, benefits in CN identification from $F0$ separation alone (by 6 ST) or onset separation alone (300 ms) were similar (approximately 24.8 percentage points). Further, when these two cues were then combined, the benefit from 6 ST and 300 ms separation was 54 percentage points; that is, doubled compared to benefits provided by each cue alone. This would demonstrate additivity of the benefits for the two cues, as well as suggesting independence of the underlying mechanisms.

Regarding group differences, *post hoc* testing showed that CN identification of the YNH group was significantly better compared to either the ENH or EHI group, and that the performance of the EHI group did not differ significantly from that of yet the ENH listeners. This pattern of group differences was clearly not the same as that observed in Experiment 1, for which it was primarily the EHI group alone that differed from the other groups.

In addition, individual differences were examined for the elderly listeners' CN identification performance (in RAU) to determine whether performance was related to call-sign detection sensitivity, as had been the case in Experiment 1. CN identification and detection sensitivity was found to be significantly and positively correlated only for one of the four conditions: the 6 ST/300 ms condition ($r = 0.53$, $p = 0.03$). In addition, among the 16 older adults in Experiment 2, there

were no significant correlations between CN identification performance and high-frequency hearing loss, age, or digit span.

In summary, as in Experiment 1, the $\Delta F0$ and onset-asynchrony cues were found to improve CN identification performance. Complete randomization of the stimulus conditions resulted in overall lower CN identification performance for a given condition, except for the baseline 0 ST/0 ms condition. In terms of CN identification performance, the effect of a 6-ST shift in $F0$ was approximately the same as a 300-ms onset asynchrony for all subject groups and the combination of both cues led to further, roughly additive, benefits. Especially when the segregation cues became less uncertain (data from Exp. 2 to Exp. 1), the EHI listeners needed stronger segregation cues to identify the target color and number words (see Fig. 7). In terms of group differences, the biggest difference between Experiments 1 and 2 was that ENH performed similarly to EHI listeners with increasing uncertainty (in Experiment 2), but both older groups were significantly worse than the YNH group.

3. CN intrusions among incorrect responses

Figure 8 plots the proportion of the CN intrusions (in RAU), relative to the total incorrect responses, measured from the first (solid) and second (striped) experiments for the YNH, ENH, and EHI groups. In Experiment 2, the mean CN intrusion collapsed across conditions was 92, 88, and 84 RAU for YNH, ENH, and EHI groups, respectively. When the intrusion rate was collapsed across groups, the CN intrusion was 109, 78, 106, and 60 RAU for 0 ST/0 ms, 0 ST/300 ms, 6 ST/0 ms, and 6 ST/300 ms conditions, respectively.

In order to investigate the effect of group and stimulus condition on CN intrusions, a 3×4 ANOVA was calculated and *post hoc* paired comparisons were performed. The results showed that CN intrusion was significantly ($p < 0.01$) affected by both group [$F(2, 21) = 7.4$] and stimulus condition [$F(3, 63) = 171.9$]. There was no significant interaction between two factors. Paired-comparison results showed that the CN intrusion rate for the 0 ST/0 ms condition did not significantly differ from the intrusion rate for the 6 ST/0 ms condition, but was significantly greater than the intrusion

rates for the other two stimulus conditions. The intrusion proportion obtained with either $\Delta F0$ of 6 ST alone or onset separation of 300 ms alone was also significantly greater compared to intrusions for the stimulus condition involving both segregation cues (6 ST/300 ms). Finally, significantly greater intrusions occurred with the $\Delta F0$ cue alone than with the onset-asynchrony cue alone. With regard to group differences in intrusion proportions, non-significant group differences in intrusions were observed between the YNH and ENH groups, as well as between the ENH and EHI groups, but the proportion of intrusions in YNH listeners was significantly higher than in EHI listeners. From the observation of fewer intrusions in EHI listeners, it would appear that the EHI subjects were more prone than YNH to make random errors, rather than intrusions, suggesting that they may not have been able to process either the target or the competing CN coordinates.

To examine whether the elderly listeners' CN intrusion and their identification performance was correlated, correlational analyses were conducted for each of the four stimulus conditions. The results showed that when the stimulus condition was 6 ST/300 ms, the condition most likely leading to maximum segregation of target and competing talkers, the CN intrusions and identification performance were significantly and negatively correlated ($r = -0.86$, $p < 0.001$) in older adults. Thus, as in Experiment 1, this may suggest that those older adults who were less able to inhibit the competing CN coordinate had poorer CN identification in the 6 ST/300 ms condition. There were no significant correlations between the intrusion proportions and the older adult's hearing thresholds, ages, or digit spans.

V. GENERAL DISCUSSION

A. The use of ΔF0 and onset separation cues for detection versus identification

The purpose of the current study was to compare the benefit from fundamental frequency and temporal onset differences between two competing sentences across groups differing in hearing and age, when they were asked to detect and identify the target words within the CRM sentence pairs. In Experiment 1, the stimulus conditions were blocked such that the elimination of uncertainty was expected to facilitate detecting or identifying the target message. In contrast, Experiment 2 was conducted with full randomization of the entire stimulus set to add target- and masker-uncertainty, which may have negatively impacted overall CN identification performance.

Across Experiments 1 and 2, several similar results were observed. First, across listener groups, no differences were observed within an experiment in the ability to detect the cue word Baron. Second, when no or little difference in $F0$ or onset existed between the CRM sentences comprising a pair, all the listeners had difficulty identifying the target-message content. Third, a greater dissimilarity between sentence pairs manipulated by $\Delta F0$ or onset differences greatly reduced the confusion with the competing information (intrusion errors) and increased identification performance. Fourth, although all the CN identification performance of the
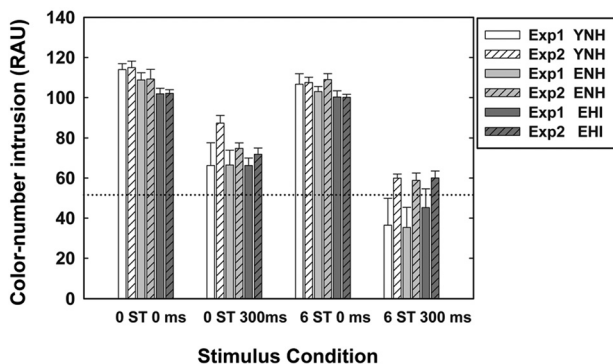

FIG. 8. Proportion of CN intrusion (in RAU) of YNH, ENH, and EHI listeners in Experiment 1 (solid) and Experiment 2 (striped) for the four types of separation cues. Line at about 52 RAU plots the proportion of intrusions obtained by chance. Abscissa labels show the $\Delta F0$ value in STs and the onset asynchrony in milliseconds (ms) for each stimulus condition.

listener groups benefited from $\Delta F0$ and onset asynchrony, the relative improvements on the identification of the target CN were smaller in older adults, especially more so in the EHI group (Experiment 1), compared to those observed in young adults.

Taken together, listener groups did not differ in detecting the cue word Baron marking the target sentence, which appeared early in the target sentence, yet differed in identifying the target CN, which was spoken in the latter part of target sentences. In our CRM task, since the detection task requires monitoring and contrasting, thus attending to the CRM utterances until the listener notices the target cue word, at least for Experiment 2, the detection performance could be relevant to the ability of listeners to divide their attention between two concurrently presented messages. As soon as the listeners identify the target sentence, then listeners need to follow the target voice and selectively attend to the CN spoken by the target voice. Thus, the potential underlying mechanism for detecting Baron would be more associated with divided attention, whereas identification performance would be associated with the ability of selective attention, assuming the target sentence had been identified earlier in processing. This general conceptualization is supported by the findings of Shafiro and Gygi (2007). In order to increase divided attention load using the CRM corpus, Shafiro and Gygi (2007) increased the number of cue words that listeners needed to detect from one (Baron) to three (Baron, "Hopper." "Tiger"), although the listeners were always required to identify the CN spoken by the talker who said Baron. With a greater load on divided attention, listeners' detection sensitivity for Baron declined, as did their CN performance.

In the current study, it appeared that older adults showed more difficulty identifying the target CN than younger adults did, even though the older adults did not differ from young adults with regard to detection of the cue word. This suggests that once the target message had been identified, older adults were substantially poorer at tracking the target voice for the remainder of the sentence, presumably due to reduced selective attention. Data for the intrusion errors support this idea. For Experiment 1, recall that at 600-ms onset asynchrony (0 ST/600 ms, 3 ST/600 ms, 6 ST/600 ms), the cue word Baron in the target sentence, which was always presented first, was presented before the later-arriving masker sentence started. As seen in Fig. 5, compared to the younger groups, the older groups (ENH, EHI) had relatively high proportions of intrusions for the 0 ST/600 ms, 3 ST/600 ms, and 6 ST/600 ms conditions. Thus, they were able to segregate the two messages but could not associate the target CN with the target message consistently. Rather, they responded with the competing CN much more so than the younger adults.

B. Effects of uncertainty

Closer inspection of the mean detection and identification performance across two experiments reveals interesting patterns. First, as seen in Fig. 6, all the listener groups had reduced detection sensitivity to Baron due to the maximum uncertainty incorporated in Experiment 2. Thus, all listener groups had more difficulty detecting the cue word under conditions of maximum uncertainty. For CN identification, however, a different pattern was observed across listener groups. For the three stimulus conditions providing segregation cues, all but the baseline reference condition (left) in Fig. 7, the trend was for the YNH and ENH groups to perform considerably worse when going from Experiment 1 to Experiment 2 whereas EHI groups performed similarly in Experiments 1 and 2, as if the EHI subjects could not receive benefit from the elimination of uncertainty in Experiment 1. In support of this notion, the lone condition for which the EHI subjects performed considerably better in Experiment 1 relative to Experiment 2 was the one which provided the strongest sound-segregation cues (6 ST/300 ms condition). The young and older NH subjects showed CN identification performance differences across experiments, with considerably better performance in Experiment 1, for all three stimulus conditions providing sound-segregation cues. Mackersie et al. (2011) and Humes and Coughlin (2009) also found significant effects of the stimulus uncertainty on CRM-task performance, regardless of listeners' hearing status or age. These observations regarding between-experiment differences across groups, however, must be tempered by the fact that in all cases different individuals comprised the YNH, ENH, and EHI groups in Experiments 1 and 2. Thus, some of the group differences across experiments could be due to different individuals comprising the various groups in each experiment.

Second, we conducted additional analyses to examine whether the detection or identification performance of listener groups was impacted by which message received the cue with higher uncertainty (Exp. 2). Recall that $F0$ or onset separation cue was applied half of the time to the target and half of the time to the masker. Statistical analysis showed that CN identification performance did not differ by which sentence had received the $F0$ or onset segregation cue. However, detection performance was significantly different for the onset-asynchrony cue [$F(1, 21) = 23.04$, $p < 0.01$], showing better d' when the target sentence came first (the mean d' was 3.0 across the three groups) compared to when the target sentence came later (the mean d' was 2.5 across the groups). This pattern was observed for both the ENH and EHI groups, but not for the YNH group, as revealed by a significant two-way interaction ($p < 0.05$) between the onset-asynchrony manipulation and group, as well as follow-up paired comparisons. Specifically, the d' for the detection of the cue word by the YNH group was similar (2.8–2.9) regardless of which one of the two sentences was presented first. In contrast, the two older groups better detected Baron when the target sentence came first (mean $d' = 3.1$–3.2) compared to when the masker sentence came first (mean $d' = 2.4$). However, as seen from Fig. 6, the trial-to-trial variability provided in Exp. 2 significantly reduced detection performance compared to Exp. 1, regardless of whether the onset asynchrony was applied to the target or the masker.

In addition, the same cue-word detection task was required across Experiments 1 and 2. Considering a possibility that the increased uncertainty in Exp. 2 would result in differential response biases or decision criteria across groups to detect the cue word Baron, we additionally estimated the

response bias (beta, β) (Green and Swets, 1966; Macmillan and Creelman, 2005). The quantity β was calculated from the equation, $\beta = \exp\{-0.05 \ [Z(\text{Hit rate})^2 + Z(\text{False alarm rate})^2]\}$ (Needleman and Crandell, 1997). Results for β in both Exp. 1 and Exp. 2 showed that mean β values were close to 1.0 throughout (ranged from 0.8–1.3), and were similar across the groups. Although we found no significant effect of uncertainty on response bias of listeners, the β values appeared to decrease slightly, but significantly, with increasing onset asynchrony, but not with $F0$ shift, for both Exp. 1 and Exp. 2.

ACKNOWLEDGMENTS

Alain, C. (**2007**). "Breaking the wave: effects of attention and learning on concurrent sound perception," Hear. Res. **229**, 225–236.

Alain, C., McDonald, K. L., Ostroff, J. M., and Schneider, B. (**2001**). "Age-related changes in detecting a mistuned harmonic," J. Acoust. Soc. Am. **109**, 2211–2216.

Alain, C., Reinke, K., He, Y., Wang, C., and Lobaugh, N. (**2005**). "Hearing two things at once: neurophysiological indices of speech segregation and identification," J. Cogn. Neurosci. **17**, 811–818.

Allen, K., Carlile, S., and Alais, D. (**2008**). "Contributions of talker characteristics and spatial location to auditory streaming," J. Acoust. Soc. Am. **123**, 1562–1570.

American National Standards Institute. (**1997**). *Method for the Calculation of the Speech Intelligibility Index*, ANSI S3.79-1997 (New York).

American National Standards Institute. (**1999**). *Maximum Permissible Ambient Levels for Audiometric Test Rooms*," ANSI S3.1-1999 (New York).

American National Standards Institute. (**2004**). *Specification for Audiometers*, ANSI S3.6-2004 (New York).

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–Prutsman, 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.

Arehart, K. H., Souza, P. E., Muralimanohar, R. K., and Miller, C. W. (**2011**). "Effects of age on concurrent vowel perception in acoustic and simulated electroacoustic hearing," J. Speech Lang. Hear. Res. **54**, 190–210.

Assmann, P. F. (**1999**). "Fundamental frequency and the intelligibility of competing voices," in Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS99), pp. 179–182.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P. F., and Summerfield, Q. (**1994**). "The contribution of waveform interactions to the perception of concurrent vowels," J. Acoust. Soc. Am. **95**, 471–484.

Backoff, P. M., and Caspary, D. M. (**1994**). "Age-related changes in auditory brainstem responses in Fischer 344 rats: Effects of rate and intensity," Hear. Res. **73**, 163–172.

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, London), pp. 263–269.

Boettcher, F. A., Mills, J. H., Swerdloff, J. L., and Holley, B. L. (**1996**). "Auditory evoked potentials in aged gerbils: Responses elicited by noises separated by a silent gap," Hear. Res. **102**, 167–178.

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

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), pp. 1–790.

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

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

Brungart, D. S., Iyer, N., and Simpson, B. D. (**2006**). "Monaural speech segregation using synthetic speech signals," J. Acoust. Soc. Am. **119**, 2327–2333.

Brungart, D. S., and Simpson, B. D. (**2004**). "Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty," J. Acoust. Soc. Am. **115**, 301–310.

Brungart, D. S., and Simpson, B. D. (**2007**). "Effects of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **122**, 1724–1734.

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.

Carroll, J., and Zeng, F. G. (**2007**). "Fundamental frequency discrimination and speech perception in noise in cochlear implant simulations," Hear. Res. **231**, 42–53.

Chalikia, M. H., and Bregman, A. S. (**1989**). "The perceptual segregation of simultaneous auditory signals: pulse train segregation and vowel segregation," Percept. Psychophys. **46**, 487–496.

Committee on Hearing, Bioacoustics and Biomechanics (CHABA) (**1988**). "Speech understanding and aging," J. Acoust. Soc. Am. **83**, 859–895.

Cooper, W. E., and Sorensen, J. M. (**1981**). *Fundamental Frequency in Sentence Production* (Springer-Verlag, New York), pp. 1–213.

Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0," J. Acoust. Soc. Am. **93**, 3454–3467.

Culling, J. F., and Darwin, C. J. (**1994**). "Perceptual and computational separation of simultaneous vowels: cues arising from low frequency beating," J. Acoust. Soc. Am. **95**, 1559–1569.

Darwin, C. J. (**1981**). "Perceptual grouping of speech components differing in fundamental frequency and onset-time," Q. J. Exp. Psychol. **33**, 185–208.

Darwin, C. J. (**2001**). "Auditory grouping and attention to speech," in Proceedings of the Institute of Acoustics, **23**, pp. 165–172, http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/papers/2001DarwinWISP.pdf (Last viewed July 2, 2010).

Darwin, C. J. (**2008**). "Listening to speech in the presence of other sounds," Philos. Trans. R. Soc. London, Ser. B **363**, 1011–1021.

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.

de Cheveigné, A. (**1997**). "Concurrent vowel identification: III. A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. **101**, 2857–2865.

Durlach, N. I., 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.

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.

Freyman, R. L., Helfer, K. S., and Balakrishnan, U. (**2007**). "Variability and uncertainty in masking by competing speech," J. Acoust. Soc. Am. **121**, 1040–1046.

Gelfer, M. P., and Mikos, V. A. (**2005**). "The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels," J. Voice **19**, 544–554.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York), pp. 1–479.

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

Hasher, L., Lustig, C., and Zacks, R. (**2007**). "Inhibitory mechanisms and the control of attention," in *Variation in Working Memory*, edited by A. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, and J. N. Towse (Oxford University Press, Oxford), pp. 227–249.

Hasher, L., and Zacks, R. T. (**1988**). "Working memory, comprehension, and aging: A review and new view," in *The Psychology of Learning and Motivation: Advances in Research and Theory*, edited by G. H. Bower (Academic Press, New York), pp. 193–225.

Hasher, L., Zacks, R. T., and May, C. P. (**1999**). "Inhibitory control, circadian arousal, and age," in *Attention and Performance XVII, Cognitive Regulation of Performance: Interaction of Theory and Application*, edited by D. Gopher and A. Koriat (MIT Press, Cambridge, MA), pp. 653–675.

Harrington, J., Palethorpe, S., and Watson, C. I. (**2007**). "Age-related changes in fundamental frequency and formants: A longitudinal study of four speakers," in *Interspeech 2007*, http://www.phonetik.unimuenchen.de/jmh/research/papers/interspeechage07.pdf (Last viewed July 2, 2009).

Hedrick, M. S., and Madix, S. G. (**2009**). "Effect of vowel identity and onset asynchrony on concurrent vowel identification," J. Speech Lang. Hear. Res. **52**, 696–705.

Helfer, K. S., Chevalier, J., and Freyman, R. L. (**2010**). "Aging, spatial cues, and single-versus dual-task performance in competing speech perception," J. Acoust. Soc. Am. **128**, 3625–3633.

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

Humes, L. E., and Coughlin, M. P. (**2009**). "Aided speech-identification performance in single- talker competition by elderly hearing-impaired listeners," Scand. J. Psychol. **50**, 485–494.

Humes, L. E., Dirks, D. D., Bell, T. S., and Kincaid, G. E. (**1987**). "Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers," J. Acoust. Soc. Am. **81**, 765–773.

Humes, L. E., Lee, J. H., and Coughlin, M. P. (**2006**). "Auditory measures of selective and divided attention in young and older adults using single-talker competition," J. Acoust. Soc. Am. **120**, 2926–2937.

Kawahara, H., Masuda–Katsuse, I., and de Cheveigné, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction," Speech Commun. **27**, 187–207.

Lentz, J. J., and Marsh, S. L. (**2006**). "The effect of hearing loss on identification of asynchronous double vowels," J. Speech Lang. Hear. Res. **49**, 1354–1367.

Loizou, P. (**2000**). "COLEA: A MATLAB Software Tool for Speech Analysis," http://www.utdallas.edu/~loizou/speech/colea.htm (Last viewed July 2, 2009).

Mackersie, C. L., Dewey, J., and Guthrie, L. A. (**2011**). "Effects of fundamental frequency and vocal-tract length cues on sentence segregation by listeners with hearing loss," J. Acoust. Soc. Am. **130**, 1006–1019.

Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, 2nd ed. (Erlbaum, Mahwah, NJ), pp. 1–512.

Meddis, R., and Hewitt, M. J. (**1992**). "Modeling the identification of concurrent vowels with different fundamental frequencies," J. Acoust. Soc. Am. **91**, 233–245.

Mills, J., Schmiedt, R. A., Schulte, B. A., and Dubno, J. R. (**2006**). "Age-related hearing loss: A loss of voltage, not hair cells," Semin. Hear. **27**, 228–236.

Murray, I. R., and Arnott, J. L. (**1993**). "Toward to simulation of emotion in synthetic speech: A review of the literature on human vocal emotion," J. Acoust. Soc. Am. **93**, 1097–1108.

Needleman, A. R., and Crandell, C. C. (**1997**). "Speech perception in noise by listeners with hearing impairment and simulated sensorineural hearing loss," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahweh, NJ), Chap. 29, pp. 461–473.

Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference," J. Acoust. Soc. Am. **125**, 457–468.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–185.

Raza, A., Milbrandt, J. C., Arneric, S. P., and Caspary, D. M. (**1994**). "Age-related changes in brainstem auditory neurotransmitters: Measures of GABA and acetylcholine function," Hear. Res. **77**, 221–230.

Rossi–Katz, J., 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., 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.

Shafiro, V., and Gygi, B. (**2007**). "Perceiving the speech of multiple concurrent talkers in a combined divided and selective attention task," J. Acoust. Soc. Am. **122**, 229–235.

Shinn–Cunningham, B. G., and Best, V. (**2008**). "Selective attention in normal and impaired hearing," Trends Amplif. **12**, 283–299.

Snyder, J. S., and Alain, C. (**2005**). "Age-related changes in neural activity associated with concurrent vowel segregation," Brain Res. Cognit. Brain Res. **24**, 492–499.

Snyder, J. S., and Alain, C. (**2007**). "Toward a neurophysiological theory of auditory stream segregation," Psychol. Bull. **133**, 780–799.

Sommers, M. S. (**1997**). "Speech perception in older adults: the importance of speech-specific cognitive abilities," J. Am. Geriatr. Soc. **45**, 633–637.

Sommers, M. S., and Danielson, S. M. (**1999**). "Inhibitory processes and spoken word recognition in young and older adults: the interaction of lexical competition and semantic context," Psychol. Aging **14**, 458–472.

Srinivasan, S., and Wang, D. (**2008**). "A model for multitalker speech perception," J. Acoust. Soc. Am. **124**, 3213–3224.

Stanislaw, H., and Todorov, N. (**1999**). "Calculation of signal detection theory measures," Behav. Res. Methods Instrum. Comput. **31**, 137–149.

Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.

Stubbs, R. J., and Summerfield, Q. (**1988**). "Evaluation of two voice-separation algorithms using normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **84**, 1236–1249.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Summerfield, Q., and Assmann, P. F. (**1989**). "Auditory enhancement and the perception of concurrent vowels," Percept. Psychophys. **45**, 529–536.

Summerfield, Q., and Assmann, P. F. (**1991**). "Perception of concurrent vowels: effects of harmonic misalignment and pitch-period asynchrony," J. Acoust. Soc. Am. **89**, 1364–1377.

Summerfield, Q., and Culling, J. F. (**1992**). "Auditory segregation of competing voices: Absence of effects of FM or AM coherence," Philos. Trans. R. Soc. London, Ser. B **336**, 357–365.

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.

Tun, P. A., O'Kane, G., and Wingfield, A. (**2002**). "Distraction by competing speech in young and older adult listeners," Psychol. Aging. **17**, 453–467.

Tun, P. A., and Wingfield, A. (**1999**). "One voice too many: Adult age differences in language processing with different types of distracting sounds," J. Gerontol. B Psychol. Sci. Soc. Sci. **54**, 317–327.

Vestergaard, M. D., Fyson, N. R., 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.

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

Wechsler, D. (**1997**). *Wechsler Adult Intelligence Scale (WISC-III)*, 3rd ed. (The Psychological Corporation, San Antonio, TX).

Wright, L. L., and Elias, J. W. (**1979**). "Age differences in the effects of perceptual noise," J. Gerontol. **34**, 704–708.