A correlational method to concurrently measure envelope and temporal fine structure weights: Effects of age, cochlear pathology, and spectral shaping[a)]

Daniel Fogerty[b)]
Department of Communication Sciences and Disorders, University of South Carolina, Columbia, South Carolina 29208

Larry E. Humes
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

The speech signal may be divided into spectral frequency-bands, each band containing temporal properties of the envelope and fine structure. This study measured the perceptual weights for the envelope and fine structure in each of three frequency bands for sentence materials in young normal-hearing listeners, older normal-hearing listeners, aided older hearing-impaired listeners, and spectrally matched young normal-hearing listeners. The availability of each acoustic property was independently varied through noisy signal extraction. Thus, the full speech stimulus was presented with noise used to mask six different auditory channels. Perceptual weights were determined by correlating a listener's performance with the signal-to-noise ratio of each acoustic property on a trial-by-trial basis. Results demonstrate that temporal fine structure perceptual weights remain stable across the four listener groups. However, a different weighting typography was observed across the listener groups for envelope cues. Results suggest that spectral shaping used to preserve the audibility of the speech stimulus may alter the allocation of perceptual resources. The relative perceptual weighting of envelope cues may also change with age. Concurrent testing of sentences repeated once on a previous day demonstrated that weighting strategies for all listener groups can change, suggesting an initial stabilization period or susceptibility to auditory training.
© 2012 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4742716]

I. INTRODUCTION

Speech intelligibility has classically been predicted by the availability and importance of the acoustical speech information in different frequency bands using Articulation-Index (AI) concepts (French and Steinberg, 1947; Fletcher and Galt, 1950), including contemporary enhancements such as the Speech Intelligibility Index (SII; ANSI, 1997). Such a method, while largely successful, is primarily based on the spectral characteristics of the signal. Modifications to account for temporal fluctuations in spectral availability may be an improvement (Rhebergen and Versfeld, 2005). However, such methods still do not account for informative temporal properties of the speech signal. Speech is inherently a temporally complex signal, containing both slow amplitude modulations of the temporal envelope (E) and fast frequency oscillations of the temporal fine structure (TFS) within each frequency band. The Speech Transmission Index (STI; Steeneken and Houtgast, 1980) incorporated many aspects of the AI framework, but focused exclusively on the preservation of the slow speech envelope. There is considerable evidence that both E and TFS temporal information contribute to speech intelligibility with the latter not included in the STI.

It is possible that the E and TFS temporal components of the speech signal contribute differently to speech intelligibility and in a frequency-dependent way. That is to say, certain temporal components may be more or less informative in different frequency bands (e.g., Apoux and Bacon, 2004; Hopkins and Moore, 2010). Defining the relative contributions of temporal components within each frequency band has the potential to refine and enhance existing models of speech intelligibility that currently consider only the distribution of spectral contributions or the E component alone. Encoding temporal contributions more fully into these models could be potentially powerful for applications to hearing aid and cochlear implant design that could predict frequency-dependent temporal contributions based upon the listener's hearing status and processing design of the device.

The current study extends recent studies investigating the relative importance of E and TFS components in three different frequency bands (Fogerty, 2011a,b). This study implements a correlational method (Berg, 1989; Richards and Zhu, 1994; Lutfi, 1995) that independently and concurrently varies the availability of these temporal components. Weighting functions obtained indicate the "importance" listeners place on suprathreshold temporal information across the broadband speech signal. Fogerty (2011a) recently demonstrated that this method is able to present the full power spectrum of speech, while using noise to independently mask E and TFS components in different frequency bands. The current study applies this method to investigating perceptual weighting strategies

---

of temporal information for young normal-hearing (YNH) listeners, older normal-hearing (ONH) listeners, older hearing-impaired (OHI) listeners with spectral shaping, and young noise-matched listeners with spectral shaping (YMC). Spectral shaping ensured audibility of the speech materials for the hearing-impaired group like a well-fit hearing aid. Comparison between the four groups thus enables the investigation of how age, cochlear pathology, and spectral shaping may impact perceptual weights.

Comparison of weighting functions is significant as it has been demonstrated that older listeners have a decreased ability to process TFS cues (Grose and Mamo, 2010; Hopkins and Moore, 2011). Age-related deficits have also been demonstrated in using temporal E information for consonant identification (Souza and Boike, 2006). Mild to moderate levels of cochlear hearing loss also reduce the ability of listeners to use TFS cues in speech (Lorenzi et al., 2006; Ardoint et al., 2010). However, listeners with hearing loss appear to have preserved use of temporal E cues for nonsense syllables when audibility of the speech signal is preserved (Turner et al., 1995). Thus, both age and hearing loss may differentially impact the processing capabilities of the listeners for these two types of temporal cues. Spectrally shaping the speech spectrum to ensure audibility also has the potential to influence how listeners use different speech information, as this processing method alters the natural *spectral* envelope of the stimulus. This has the potential to influence how cues may be compared and integrated across frequency bands. As spectral shaping is commonly used to ensure audibility of speech for listeners with hearing loss, as in conventional hearing aids, this study investigated the impact of this spectral shaping on temporal weighting functions.

It has become fairly well established that the spectral distribution of perceptual weights is dependent upon hearing status (Calandruccio and Doherty, 2008; Doherty and Lutfi, 1996, 1999; Hedrick and Younger, 2007; Mehr et al., 2001; Pittman and Stelmachowicz, 2000). Hearing-impaired listeners demonstrate different configurations of relative weights compared to normal-hearing listeners (e.g., Pittman and Stelmachowicz, 2000) and listeners with cochlear implants demonstrate great variability among listeners (Mehr et al., 2001). However, differences in spectral weights obtained for hearing-impaired listeners may reflect differences in presentation levels used to ensure audibility (Leibold et al., 2006). Thus, it may be that amplification procedures influence perceptual weights more so than the actual hearing loss after audibility of the signal is restored. Thus, the current study included a young normal hearing control group that received the same spectrally-shaped stimuli as the hearing impaired group to determine if differences are due to the underlying cochlear pathology, or the stimulus manipulation of spectral components and resulting presentation level differences.

Overall, the purpose of the current investigation was to determine the spectral distribution of perceptual weights for the temporal envelope and fine structure and how this distribution is influenced by age, cochlear pathology, and spectral shaping. Currently, little is known about how E and TFS components combine to contribute to spectral weighting patterns of listeners, and in particular, how characteristics of the listener influence these patterns. Experiment 1 was designed to examine how these listener groups use individual frequency band cues and temporal component cues (i.e., E and TFS) when the remaining speech signal is significantly masked by noise. Experiment 2 then examined how these same listeners weighed temporal component cues across frequency bands.

II. EXPERIMENT 1A: FREQUENCY BAND CONTRIBUTIONS

A. Listeners

Four groups of listeners participated in testing. Eight YNH listeners (range = 20–23 yrs, M = 21 yrs), nine ONH listeners (range = 62–83 yrs, M = 71 yrs), nine OHI listeners with spectral shaping (range = 66–84 yrs, M = 74 yrs), and nine young spectrally matched normal-hearing listeners (YMC) (range = 18–22 yrs, M = 21 yrs). Audibility for the OHI group was ensured by spectrally shaping the stimuli to be at least 15 dB above thresholds through at least 4 kHz. According to the SII, maximum speech intelligibility is obtained when the speech spectrum is 15 dB above the listener's hearing threshold, which this spectral shaping procedure ensured. This amplification target is also characteristic of a common clinically prescriptive amplification procedure, the Desired Sensation Level approach (Seewald et al., 1993). Each listener in the young spectrally matched group was randomly assigned to receive stimuli spectrally shaped as they were for one of the OHI listeners. This effectively matched the presentation level between these groups. Figure 1 displays the long-term average speech spectrum for the natural speech materials used here (bold black line) along with the average shaped spectrum presented to the OHI listeners (bold gray line).

Criteria for normal hearing consisted of audiometric thresholds ≤20 dB hearing level (HL) for young listeners
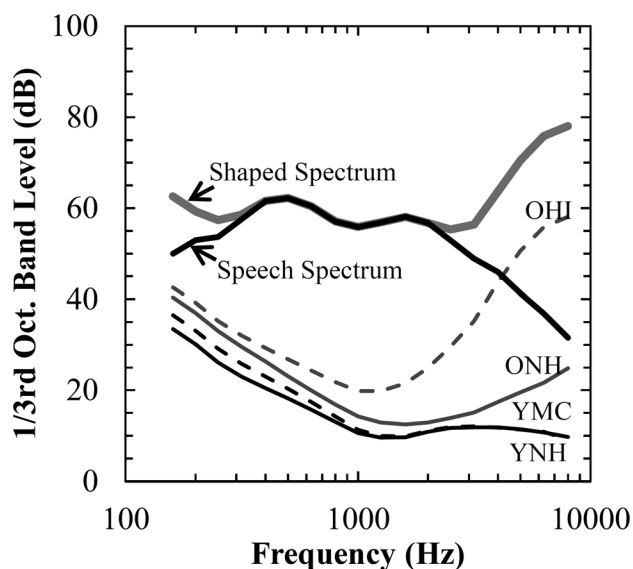

FIG. 1. Speech spectrum measured in 1/3 octave bands for the IEEE sentences presented at 70 dB SPL (bold black line) and after spectral shaping for the average hearing-impaired listener (bold gray line). Thin solid (normal-hearing) and dashed (spectrally shaped) lines display mean hearing thresholds in the test ear for the four listener groups.

and ≤ 30 dB HL for older listeners at octave frequencies from 0.25–4 kHz. Figure 1 displays the average audiometric thresholds for each of the four listener groups (thin lines). All listeners had normal tympanograms and no evidence of middle ear pathology. Other selection criteria required older listeners to obtain a score greater than 25 on the Mini Mental State Exam (Folstein *et al.*, 1975), a brief survey of cognitive status.

B. Stimuli and signal processing

IEEE/Harvard sentences (IEEE, 1969) recorded by a male talker (Loizou, 2007) were used as stimuli. These stimuli are all meaningful sentences that contain five keywords, such as in the experimental sentence (keywords in italics), "The *birch canoe slid* on the *smooth planks*." All experimental sentences were down-sampled to a sampling rate of 16 000 Hz and passed through a bank of bandpass, linear-phase, finite-impulse-response filters to process speech into three different frequency bands: 80–528 Hz (band 1), 528–1941 Hz (band 2), and 1941–6400 Hz (band 3). Frequency bands roughly corresponded to prosodic, sonorant, and obstruent linguistic classifications. As with Hopkins *et al.* (2008), each filter was chosen so that its frequency response was −6 dB at the point that the response intersected with the response of adjacent filters.

These frequency bands represented an equal distance along the cochlea using a cochlear map (Liberman, 1982) and had a similar number of equivalent rectangular bands (Moore and Glasberg, 1983). Thus, it was expected that all frequency bands would be weighted similarly.

After passing the speech stimulus through the bank of three bandpass filters, a constant noise matching the power spectrum of the speech band was added to the two masked bands to yield a signal-to-noise ratio (SNR) of −5 dB in each band. For the remaining (target) speech band, the SNR was varied individually over a range of four SNRs (17, 11, 5, −1 dB), established on the basis of prior testing. This same SNR range was used for all listeners. Thus, performance on this task varied as a function of SNR in the target frequency band. The presence of masking noise was expected to effectively elevate hearing thresholds of all listeners, thus performance was tested at similar sensation levels across the four listener groups. Figure 2 displays the spectrum of an example sentence and the matching masking noise for the condition in which the target band was the mid-frequency band presented at 11 dB SNR.

All three frequency bands were investigated, creating a total of 12 conditions (3 frequency bands × 4 SNR levels). Fifteen sentences were presented per condition with no sentences being repeated for a given listener. The 180 sentences were presented in a fully randomized order. After processing, all stimuli were up-sampled to 44 828 Hz for presentation through Tucker-Davis Technologies (TDT) System-III hardware.

C. Calibration

Stimuli were presented via TDT System-III hardware using 16-bit resolution at a sampling frequency of 48 828 Hz.
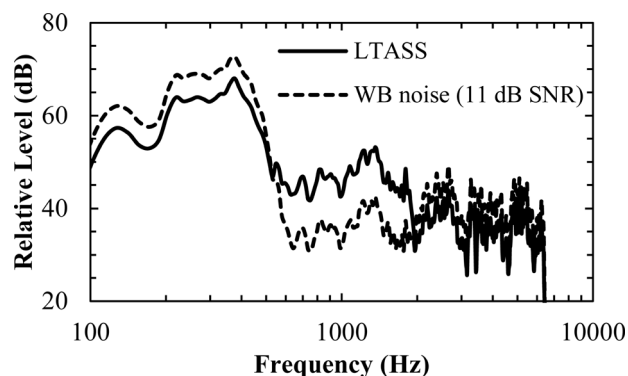

FIG. 2. Example stimulus spectrum displayed for the mid-frequency target band in the sentence, The birch canoe slid on the smooth planks. The full speech spectrum was presented (solid) with noise masking the target band at 11 dB SNR and non-target bands at −5 dB SNR (dashed). The level of the target band masking noise was varied during the experiment while all other stimulus parameters remained constant.

The output of the TDT D/A converter was passed through a headphone buffer (HB-7) and then to an ER-3 A insert earphone for monaural presentation. A calibration noise matching the long-term average speech spectrum for the sentences was set to 70 dB sound pressure level (SPL) through the insert earphone in a 2-cm^3 coupler using a Larson Davis Model 2800 sound level meter with linear weighting. Therefore, the original unprocessed wideband sentences were calibrated to be presented at 70 dB SPL. However, after filtering and noise masking, the overall sound level of the combined stimulus varied according to the individual condition. This ensured that levels representative of typical conversational level (70 dB SPL) were maintained for the individual spectral components of speech across all sentences. The overall presentation level for listeners receiving spectral shaping of the stimuli was estimated at 82 dB SPL (SD = 2 dB). Output levels were estimated by measuring one-third octave bands of the calibration noise using the sound level meter, adjusting band output by the spectral shaping provided to each individual, and calculating the summed output across all bands.

D. Procedure

All listeners completed familiarization trials of sentences selected from male talkers in the TIMIT database (Garofolo *et al.*, 1990) and processed according to the stimulus processing previously described for the experimental materials. All listeners completed experimental testing regardless of performance on the familiarization tasks. No feedback was provided during familiarization or testing to avoid explicit learning across the many stimulus trials. All listeners received different randomizations of the 180 experimental sentences (60 sentences/target band). Each trial presented a sentence preserving one target band, with the remaining two bands masked at −5 dB SNR. Each sentence was presented individually and the listener was prompted to repeat the sentence aloud as accurately as possible. Listeners were encouraged to guess. All listener responses were digitally recorded for later analysis. Only keywords repeated exactly were scored as correct (e.g., no missing or extra suffixes). In addition, keywords were allowed to be repeated

back in any order to be counted as correct repetitions. Each keyword was marked as 0 or 1, corresponding to an incorrect or correct response, respectively. Two native speakers of American English were trained to serve as raters and scored all recorded responses. For a subset of these materials, inter-rater agreement was at 97% ($r = 0.92$).

E. Results and discussion

In order to stabilize the error variance, all percent-correct keyword scores here and elsewhere were transformed to rationalized arcsine units (Studebaker, 1985) prior to analysis. The transformed percent-correct scores for the three frequency bands at four different SNRs were entered as repeated-measures variables and listener group as a between-subject variable in a general linear model analysis. A main effect of band [$F(2,62) = 4.5$, $p < 0.05$] and SNR [$F(3,93) = 524.2$, $p < 0.001$] were obtained. Interactions between band and group [$F(6,62) = 1148.3$, $p < 0.001$], as well as band and SNR [$F(6,186) = 853.4$, $p < 0.001$] were significant along with the three-way interaction of band, SNR, and group [$F(18,186) = 124.5$, $p < 0.05$]. Plotted in Fig. 3 are the mean results for each of the four groups for the low-, mid-, and high-frequency bands. As can be observed, group differences were mainly limited to the high-frequency band. However *post hoc* comparisons were conducted to fully examine group and band differences.

Independent-samples *t*-tests were conducted to investigate group age and spectral shaping differences across SNRs for each of the three bands. After Bonferroni-correction for 12 multiple comparisons (i.e., alpha = 0.05/12), results demonstrated a significant difference of age only in the third frequency band at the most favorable SNRs. Young listeners had better performance than older listeners at 11 dB [$t(33) = 4.3$, $p < 0.05$] and 17 dB SNR [$t(33) = 3.5$, $p < 0.05$]. No other group comparisons were significant ($p > 0.05$).

Differences between bands were also investigated using paired-samples *t*-tests with Bonferroni correction for 12 multiple comparisons. As only an age effect was observed, analyses were conducted separately for the pooled younger groups and pooled older groups. Young listeners scored significantly poorer for band 1 at the highest SNR compared to performance on band 2 [$t(16) = 4.6$, $p < 0.05$] and band 3 [$t(16) = 7.1$, $p < 0.05$] at that same SNR. Older listeners scored lower with the highest frequency band compared to band 1 at 11 dB [$t(17) = 9.4$, $p < 0.05$] and band 2 at 11 dB [$t(17) = 5.4$, $p < 0.05$] and 17 dB [$t(17) = 3.3$, $p < 0.05$].

Band 1 scores were also greater than band 2 at -1 dB [$t(17) = 4.3$, $p < 0.05$] and lower at 17 dB SNR [$t(17) = 3.7$, $p < 0.05$].

Overall, results suggest very similar results for all four listener groups using these frequency bands. Significant group differences were limited to the highest frequency band at the best SNRs. Under these conditions, young listeners performed better than older listeners, regardless of spectral shaping. Differences in performance between the frequency bands demonstrated poorer performance for the low-frequency band at the best SNR for young and older listeners. In addition, older listeners had a poorer performance in band 3 at the highest SNRs compared to their performance in the other two frequency bands. Overall, young and older listeners had similar sentence intelligibility scores across bands and SNR, except that older listeners had a poorer performance using high-frequency cues at the best SNRs.

III. EXPERIMENT 1B: TEMPORAL COMPONENT CONTRIBUTIONS

A. Methods

Listeners and testing procedures were the same as those in Experiment 1 A. Signal processing of the experimental sentences was conducted following the procedures described by Fogerty (2011a) and is detailed below. No sentences were repeated from Experiment 1 A.

For the creation of one trial sentence, two noise masked copies of that sentence were created at different SNRs. One copy provided the masked temporal component (i.e., E or TFS), and the other copy provided the complementary target temporal component (i.e., TFS or E). These two temporal components were from the same sentence, but masked at different SNRs. Final processing combined these two copies to result in the final stimulus.

Processing began by creating a speech-shaped noise matching the long-term average of the trial sentence. The masked temporal component was created by adding the speech-shaped noise to the trial sentence at -5 dB SNR. Then, the resulting masked sentence was passed through a bank of three analysis filters (i.e., the same as used in Experiment 1 A). The Hilbert transform extracted the E (E_mask) and TFS (TFS_mask) components within each frequency band from this masked signal (i.e., speech at -5 dB SNR).

The target temporal component was created by scaling the speech shaped noise over the range of SNR values used in Experiment 1 A (17, 11, 5, -1 dB) and added to the
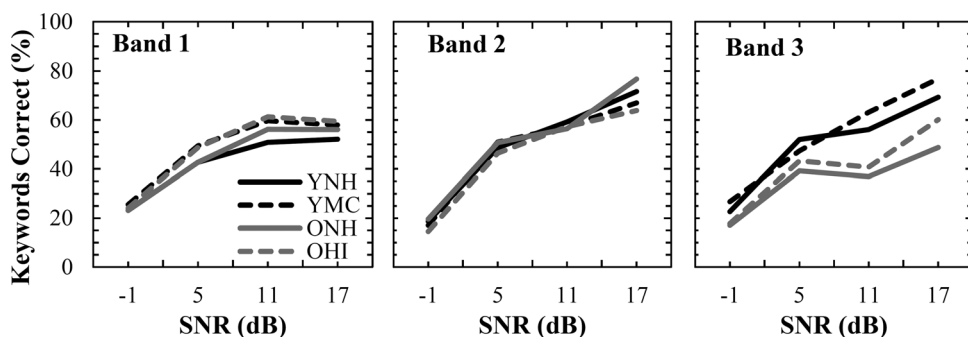

FIG. 3. Performance functions are displayed for each of the four listener groups according to the three experimental frequency bands tested.

original sentence, from which the target E (E_target) and target TFS (TFS_target) were extracted. For envelope testing, in each analysis band the E_target and TFS_mask components were combined and summed across frequency bands. The same was performed for TFS testing, where the E in each band was replaced by the E_mask for that band and combined with the TFS_target. Thus, for both E and TFS testing, the non-test temporal property was masked at a constant -5 dB SNR while the target portion was varied over the range of test SNRs. Finally, the entire stimulus was re-filtered at 6400 Hz and up-sampled to 48 828 Hz to produce the final stimulus. Thus, the final stimulus actually reconstructed the full speech spectrum, but had noise differentially added to the E and TFS components. There were a total of 8 conditions (2 temporal properties × 4 SNRs). Fifteen sentences were presented per condition for a total of 120 sentences (600 keywords).

B. Results and discussion

As in Experiment 1A, a general linear model analysis was used to investigate the temporal component (two levels) and SNR (four levels) as repeated-measures variables with group as a between-subject variable. Results demonstrated main effects of the temporal component $[F(1,31) = 196.1, p < 0.001]$ and SNR $[F(3,93) = 547.4, p < 0.001]$. An interaction between these two variables was also obtained $[F(3,93) = 10.3, p < 0.001]$. No main effect or interactions with the listener group were obtained. These findings are apparent in Fig. 4, which displays performance for E and TFS components across SNR. Little difference between listener groups is observed. Therefore, all listeners were pooled in subsequent analyses. Comparison between E and TFS components at each of the four SNRs demonstrated a significantly better performance for listeners on the E conditions for all SNRs tested ($p < 0.01$).

It is interesting to note that for these stimuli, no effect of age, hearing loss, or spectral shaping was observed for either E or TFS stimuli. This is in sharp contrast to the age differences obtained in Experiment 1A and the documented declines in TFS processing with age (e.g., Lorenzi et al., 2006). It may be that access to the full power spectrum, regardless of the noise masking, enabled listeners to maintain performance for these conditions. Finally, performance appears to level off at
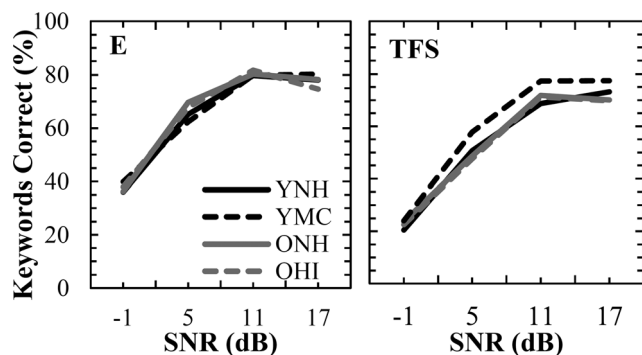

FIG. 4. Performance functions are displayed for each of the four listener groups according to the two experimental temporal components tested. E = envelope; TFS = temporal fine structure.

about 80% correct, consistent with previous investigations of E and TFS processing using three bands (Fogerty, 2011a; Shannon et al., 1995).

IV. EXPERIMENT 2: TEMPORAL COMPONENT CONTRIBUTIONS ACROSS FREQUENCY

A. Methods

Listeners and procedures were the same as those in Experiment 1. Testing for Experiment 2 occurred on a separate day. Two sets of 300 sentences were used. The first sentence set was the 300 sentences listeners heard in Experiments 1 A and 1B. Listeners only heard these sentences presented once prior to Experiment 2. These sentences were presented to examine the effect of prior stimulus exposure. The second sentence set were novel sentences unfamiliar to the listeners. All 600 sentences were randomly presented together to the listeners with the stipulation that equal numbers of set 1 and set 2 sentences occurred in the first half and second half of testing.

Signal processing combined the basic procedures implemented in Experiments 1 A and 1B to investigate E and TFS components in each of the three frequency bands. This resulted in six acoustic "channels" (two temporal properties in each of three frequency bands). These channels will be referred to according to the temporal property and band number, as in E1 (i.e., E modulation in band 1). The SNR was varied independently in each of these six different acoustic channels. As in Experiment 1, the noise matched the power spectrum of the original sentence. This ensured that the same SNR was maintained, on average, across all frequency components for each sentence. SNR values ranged from -7 to 5 dB in 3 dB steps, resulting in five different levels (i.e., -7, -4, -1, 2, and 5 dB). A correlational method (Lutfi, 1995; Richards and Zhu, 1994) was used in this experiment which correlates the SNR value of each condition on a given trial to the response accuracy of the listener on that trial. In order to use this analysis method, the SNRs in each of the six acoustic channels were independently and randomly assigned. Therefore, on a given trial the same SNR level could be presented to more than one channel (i.e., a given trial could consist of SNRs presented at -1, 2, -1, -7, 5, and 2 dB distributed across the six channels). Each SNR was presented in each acoustic channel a total of 120 trials (120 × 5 SNRs = 600 total trials). Each sentence contained 5 keywords, resulting in 3000 keywords scored. Conditions were equally distributed over the two word lists.

The speech stimulus and matching speech-shaped noise were passed through a bank of bandpass filters and the noise level was scaled according to the SNR condition for that trial. The Hilbert transform then divided each combined speech and noise band into the E and TFS. For example, if the SNR was -1 and -7 dB for the E and TFS in band 1, respectively, the Hilbert transform was used to extract the E from the -1 dB SNR band copy and the TFS from the -7 dB band copy. The noisy Hilbert components, masked at different SNRs, were then recombined. The final, full stimulus was re-filtered at 6400 Hz and up-sampled to 48 828 Hz for presentation through the TDT System-III hardware. The

final, recombined stimulus contained the full speech spectrum, with each channel differentially masked by noise according to the randomly assigned condition. As in Experiment 1, listeners' responses were audio recorded for later keyword scoring by the trained raters.

B. Results and Discussion

1. Keyword performance

Average keyword percent-correct performance across all conditions for repeated and non-repeated sentences was compared for all four listener groups. A mixed-model analysis of variance (ANOVA), with repetition as the repeated-measures variable and group as the between-subjects variable, demonstrated no main effect of repetition or group and no interaction, suggesting that repetition of sentences did not influence overall performance. Figure 5 displays the average performance for each listener group for these two sentence types (see inset). Average percent-correct performance was also broken down according to SNR for the non-repeated sentences and is displayed in the main panel of Fig. 5. SNR was entered as a repeated-measure variable along with group as a between-subject variable into an ANOVA to investigate group differences as a function of SNR. No main effect of group or interaction was obtained for these data, indicating similar performance functions for the four groups as well.

2. Correlational weights

All listener groups achieved similar overall keyword percent-correct scores for both repeated and non-repeated sentences. However, it may be that these listener groups achieve these scores by significantly different means. That is, different listener groups may place different perceptual weight on each channel, possibly emphasizing cues that are more salient or de-emphasizing cues that are more difficult

to process. Toward this end, perceptual weights were calculated for these listener groups.

Results for repeated (list 1) and non-repeated (list 2) words were analyzed separately. To determine the relative weighting of each acoustic channel, a point-biserial correlation was calculated trial-by-trial between the individual word score and the SNR in each channel for that trial. Each correlation was calculated over 1500 points (300 sentences × 5 keywords), for a total of 12 weights (i.e., E and TFS across the three bands for each of the two word lists).

Correlations for the two word lists were normalized to 1.0 to reflect the relative weight assigned to each channel and are plotted in Fig. 6. Results for each of the two word lists are analyzed separately below.

a. Non-repeated, novel sentences. A mixed-model ANOVA was conducted with temporal component (E, TFS) and frequency band (3 bands) repeated-measures factors and the listener group (YNH, ONH, OHI, and YMC) as a between-subject factor. The main effects for temporal component [$F(1,31) = 584.8$, $p < 0.001$] and band [$F(2,62) = 12.8$, $p < 0.001$] were significant, with no significant main effect of group. However, group significantly interacted with band [$F(6,62) = 6.3$, $p < 0.001$]. Temporal and frequency band weights also interacted [$F(2,62) = 6.8$, $p = 0.002$].

Bonferroni-corrected *post hoc* tests were conducted to further investigate main effects and interactions. Table I displays the results for paired *t*-tests between the different acoustic channel weights for the four listener groups. Overall, within-group comparisons demonstrated that E cues were weighted more than TFS cues across frequency bands for all listener groups (significantly for bands 1 and 3), ONH listeners weighted E2 more than the other E channels, and older listeners also placed the most weight on TFS2 compared to the other two TFS channels. Therefore, E and mid-frequency cues appear to receive the most weight, particularly for older listeners. This is in agreement with previous results for a separate group of YNH listeners (Fogerty, 2011a).

Independent sample *t*-tests were also conducted to compare relative weights between the four listener groups. Significant differences were largely limited to E cues, with all listener groups weighting TFS cues similarly across frequency. No significant difference was obtained between ONH and OHI listeners, although higher E2 weights were apparent for the ONH listeners. Thus, cochlear pathology did not appear to influence weights after audibility of the speech materials was restored for the OHI group. To investigate the effect of age, the two young and two older groups were pooled. Pooled comparisons indicated that older listeners weigh E2 more [$t(33) = -3.1$, $p = 0.004$] and E3 less [$t(33) = 3.3$, $p = 0.002$] than young listeners. This was most evident for the ONH listeners. Pooled comparisons were also investigated to examine the effect of spectral shaping. Comparisons indicated that the groups with spectral shaping had higher E1 weights [$t(33) = 2.9$, $p = 0.006$] and reduced E2 weights [$t(33) = 3.0$, $p = 0.005$] compared to the groups without spectral shaping.

These results suggest a combined effect of age and of spectral shaping. Age appears to reduce the weighting of
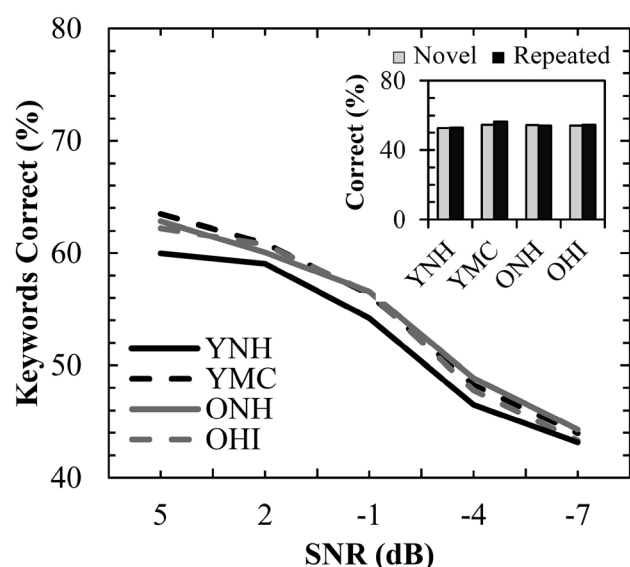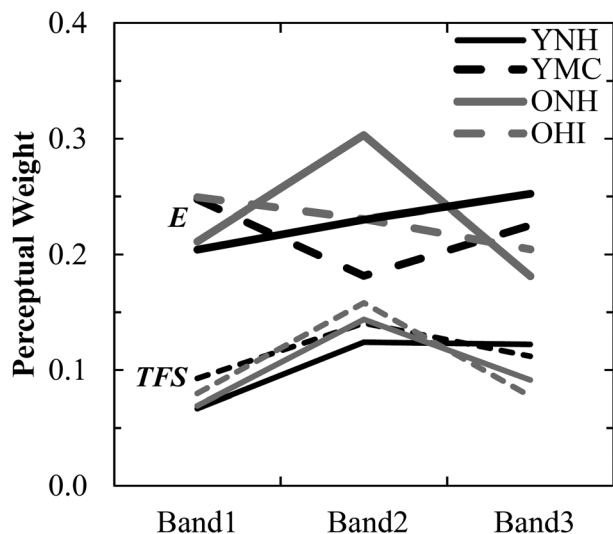

FIG. 5. Mean keyword accuracy averaged across all experimental conditions at each SNR tested. Inset displays average performance collapsed across SNR for novel (gray) and repeated (black) sentences.
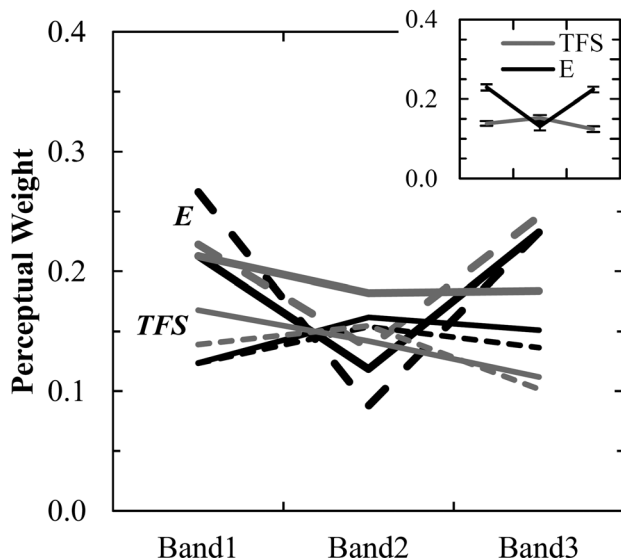
a) Not repeated

b) Repeated

FIG. 6. Relative perceptual weights obtained for the four listener groups by using a point-biserial correlational analysis across the six acoustic channels for keyword scoring. (a) Weights obtained for novel sentences unfamiliar to the listeners, (b) weights obtained for sentences heard by listeners once previously during Experiment 1 on a different testing day. Repeated sentences were randomly interspersed with the novel sentences from (a) during testing. Bold lines display E weights; thin lines display TFS weights. Inset in (b) displays the mean relative weights averaged across all listener groups, as only one group comparison (E3 between ONH and OHI listeners) resulted in significant differences for the repeated sentences.

temporal amplitude cues in the high frequencies, possibly due to the faster temporal modulation rates conveyed in those bands (Greenberg *et al.*, 1998). Spectral-shaping, which emphasized the high frequencies, appeared to increase the perceptual weighting of temporal modulations in the low-frequency band. Even though the different listener groups had some differences in E weights, TFS cues were weighted similarly and may represent stable perceptual cues that are available for all listeners across different contexts.

TABLE I. Results of paired-comparison *t*-tests between channel weights for each of the four listener groups for non-repeated sentences.

	YNH	YMC	ONH	OHI
E1 vs TFS1	17.3*	8.7*	9.5*	6.8*
E2 vs TFS2	13.7*		5.6*	
E3 vs TFS3	7.4*	4.9*	6.3*	6.0*
E1 vs E2				
E1 vs E3				
E2 vs E3			4.2	
TFS1 vs TFS2			−5.5*	−5.4*
TFS1 vs TFS3				
TFS2 vs TFS3				3.9

$p < 0.005$; *$p < 0.001$

This may be true even though older listeners may have reduced processing of these TFS cues (Lorenzi *et al.*, 2006). Finally, it is important to note that the various groups of listeners achieved similar keyword correct scores across SNR. Thus, while actual measured performance was similar across the four groups, these results suggest that older listeners and listeners with amplification may be "listening" to speech in a different way than YNH listeners. These differences, evident in the perceptual weights obtained in this study, may become more evident in more adverse listening conditions that tax sensory and/or cognitive processing of the stimulus or in tasks that require processing of specific acoustic channels. Alternatively, differences in perceptual weighting may have resulted in different types of errors while maintaining overall performance, although preliminary analyses have not revealed such error patterns.

b. Repeated sentences. A second ANOVA investigated the influence of the same factors as before on the perceptual weights, this time for the repeated sentences. Main effects for temporal component [$F(1,31) = 108.1$, $p < 0.001$] and band [$F(2,62) = 11.5$, $p < 0.001$] were significant, with no significant main effect of group. However, group significantly interacted with band [$F(6,62) = 2.3$, $p < 0.05$]. Temporal and frequency-band weights also interacted [$F(2,62) = 56.4$, $p < 0.001$]. These results largely parallel findings for the non-repeated sentences.

Paired *t*-tests were conducted for each listener group for comparison of weights between channels. These weights were consistent with results for non-repeated sentences and are summarized here and in Table II. All listener groups significantly weighted E more than TFS in band 1 (except YNH and ONH) and band 3. No difference between E2 and TFS2 was observed. E2 was weighted less than E3 by YNH, YMC, and OHI listeners. OHI had a higher weight for TFS2 compared to TFS3, and YMC weighted E1 more than E2. Thus, as observed in Fig. 6, differences in weights for E1–E3 and TFS1–TFS3 were reduced after prior exposure to the sentences.

Independent-sample *t*-tests were again conducted between groups. The only group comparison that reached significance was that OHI had higher E3 weight than ONH [$t(16) = 3.1$, $p < 0.008$]. Overall, all listener groups overwhelmingly weighted the six acoustic channels similarly. This is of note

TABLE II. Results of paired-comparison t-tests between channel weights for each of the four listener groups for repeated sentences.

	YNH	YMC	ONH	OHI
E1 vs TFS1		13.0*		4.4
E2 vs TFS2				
E3 vs TFS3	6.9*	4.5	6.7*	12.0*
E1 vs E2		5.8*		
E1 vs E3				
E2 vs E3	−5.6*	−6.2*		−5.6*
TFS1 vs TFS2				
TFS1 vs TFS3				
TFS2 vs TFS3				4.1

$p < 0.005$; *$p < 0.001$

given the significant group differences for the non-repeated sentences. Even a single repetition of the sentences may facilitate similar weighting strategies among these listener groups. Given the similarity in group performance, the inset of Fig. 6(b) displays the mean weighting function collapsed across groups. This function may represent the "typical" and stabilized relative weighting profile of the average listener for familiar, everyday listening situations.

c. Comparison between novel and repeated sentences. Already evident in Fig. 6 is the fundamental difference in weights obtained for the repeated sentences. This finding is remarkable given the large number of sentences presented in randomized, highly degraded conditions that were interspersed with 300 additional novel sentences. Furthermore, listeners only heard these sentences presented once before on a separate day. The change in perceptual weights for these repeated sentences is not due to talker familiarity or procedural learning, as they were randomized along with the non-repeated sentences. While it is possible that some of the sentences were familiar to the listeners, the large number of sentences and high level of signal degradation during the original presentation would limit the sentence recall ability of all listeners, young and old alike. These results suggest that perceptual weights are highly plastic, susceptible to acoustic changes in the environment (such as during temporal interruption, see Fogerty, 2011b), or top-down cognitive influences due to familiarity. Thus, these perceptual weights may represent successful, malleable targets for auditory training.

To investigate which channels were most affected by repetition, weights for repeated and non-repeated sentences were compared for each listener group. Paired t-tests with Bonferroni correction demonstrated a reduction of E2 weight given a second presentation for all listener groups except YMC ($p < 0.05$). Older listener groups also increased weight to TFS1 on the second presentation [ONH: $t(8) = −11.0$, $p < 0.001$; OHI: $t(8) = −4.5$, $p < 0.008$]. No other comparisons reached significance.

V. GENERAL DISCUSSION

A. Correlations among older listeners

A correlational analysis was conducted on results for the older listeners to determine potential factors that may be related to the obtained perceptual weights. These additional factors included perceptual weights that were derived from results for the individual frequency bands (Experiment 1A) and temporal components (Experiment 1B) based on the trial-by-trial performance and the trial SNR. In addition, all listeners also completed speech-in-noise testing for monosyllabic words and TIMIT sentences (Garofolo *et al.*, 1990) at 0 dB SNR. Correlations of these measures, mean keyword accuracy, age, and high-frequency pure-tone averages (1, 2, 4 kHz) with the relative weights for non-repeated sentences from Experiment 2 were measured.

A summary of significant findings is provided in Table III. Weights for each of the three independent frequency bands were positively correlated with the E for that respective band. In addition, band 2 was negatively correlated with E1, indicating that listeners who were more influenced by band 2 cues placed less weight on low-frequency E cues. Interestingly, band weights did not correlate with TFS cues. In addition, perceptual weights derived from the independent E and TFS testing in Experiment 1B did not significantly correlate with the relative weights obtained in the combined condition.

Age was significantly correlated with E2, as expected from the observed group differences. However, pure-tone averages were not significantly correlated with perceptual weights. Finally, correlations with speech-in-noise measures were obtained only with mean keyword performance (words: $r = 0.53$; sentences: $r = 0.58$). These measures were not significantly correlated with the relative weights obtained here for either independent (Experiment 1) or concurrent (Experiment 2) testing.

Combined, these results suggest that processing of temporal information within frequency bands is mostly related to extracting E cues from the signal. Therefore, temporal information contributing to spectral band importance in speech intelligibility models, such as the SII, may be mostly influenced by the preservation of E cues with TFS cues serving a more minor role. This also highlights the potential importance of models such as the STI, which consider temporal envelope modulation. However, further systematic work is required to investigate this possibility.

The perceptual weighting of these cues are also more susceptible to age-related suprathreshold factors. Even though older listeners had reduced E3 weights, small elevations in hearing thresholds were not correlated with perceptual weights. In addition, while speech-in-noise perceptual abilities are associated with overall performance, for these

TABLE III. Pearson correlation coefficients for older listeners between relative perceptual weights obtained under individual testing of frequency bands in Experiment 1A and concurrent testing of acoustic channels in Experiment 2. Age is also included as a factor.

	E1	E2	E3
Band 1	0.54		
Band 2	−0.65*	0.76*	
Band 3			0.65*
Age	−0.49	0.49	

$p < 0.05$, *$p < 0.01$

stimuli, overall performance was not associated with how listeners perceptually weighted these acoustic channels. Recall that this was also the case for the comparison of repeated and non-repeated sentences from Experiment 2. That is, different weighting profiles were observed for these two sentence sets but overall performance for these two sentence sets did not differ at each SNR. These results call into question the importance of relative weights in determining speech recognition performance on these tasks as different weights resulted in the same overall performance. However, individual differences in perceptual weights may be more associated with speech-in-noise abilities for more complex noise conditions, such as during fluctuating or interrupting noise, which may more directly interfere with E cues in the signal. That is, weighting differences may result in performance differences for adverse listening conditions that directly interfere with temporal speech properties. The association between overall speech performance and an individual's reliance on certain acoustic channels represents an important area of investigation as training or signal enhancement targeting such cues may be potential interventions to assist a listener's performance under such adverse listening conditions. Given the plasticity of perceptual weights under limited familiarization, such training may be particularly successful if a link is established between overall perceptual performance and perceptual weights. However, the stability of relative perceptual weights within the same listeners needs to be identified, particularly over larger sets of stimuli. The stability of perceptual weights has largely been overlooked in studies employing similar correlational methods (although see Doherty and Turner, 1996).

B. Discriminant analysis

The listeners were pre-classified in this study according to age and hearing status. This study has reviewed differences in perceptual weighting for these listener groups. The relationship of these perceptual weights to these listener groups can be further established if it is possible to classify listeners according to their weighting configurations. If so, then these weights could be used to predict properties of the listener or vice versa. Toward this end, two stepwise discriminant analyses were conducted; first to predict age and second to predict whether the listener received spectral shaping. For age, E3, TFS1, TFS3 were significant predictors $[F(3,31) = 10.1, p < 0.001, \lambda = 0.5]$. The leave-one-out analysis (i.e., predicting class membership of one case on the basis of all other cases) predicted 74% of cases. Two-thirds of errors occurred from misclassifying older listeners as young with the majority of these misclassified cases being OHI listeners who received spectral shaping. For spectral shaping, E2 and TFS3 were found to be predictors $[F(2,32) = 7.2, p < 0.01, \lambda = 0.7]$. The leave-one-out analysis again predicted 74% of cases. Misclassifications occurred equally for young and older listeners.

These results are consistent with the group comparisons. First, both analyses highlight the influence of age on perceptual weighting of the high-frequency band. Older listeners weighted this band less than young listeners, which signifi-

cantly classified listeners. Note that E2, although significantly higher for older listeners, was not a significant predictor of age. Second, E2 was found to be more associated with spectral shaping, which was related to reduced weighting. However, the discriminant analysis also reveals sensitivity of TFS weights in predicting age and spectral shaping. While all listeners equally weighted TFS cues, these weights did add to the classification scores in the stepwise discriminant analysis, particularly for classifying listeners who received spectral shaping.

C. Effect of sentence repetition

Results demonstrated a marked change in perceptual weighting, particularly for E2, for sentences repeated a second time. This change could not be accounted for by procedural learning factors, as repeated sentences were randomly interleaved with novel sentences. Fogerty (2011a) also measured perceptual weights for these channels with a separate group of YNH listeners (18–26 yrs). Importantly, these listeners were tested on twice as many novel sentences as presented here and no repeated sentences were presented. Comparison to these previous results could help determine how stable these novel perceptual weights are and if additional procedural experience would alter perceptual weights. Figure 7 displays weights for YNH listeners to repeated and novel sentences, as well as the separate novel sentence testing by Fogerty (2011a). The novel testing here and by Fogerty (2011a) demonstrates very similar trends. The most notable difference is a reduction of E2 relative weight for the repeated sentence testing.

Importantly, many of the group differences were eliminated for the repeated sentences, possibly reflecting a convergence on the typical weighting function in familiar, everyday listening environments. This average function is displayed in the Fig. 6(b) inset, collapsing across the repeated sentence
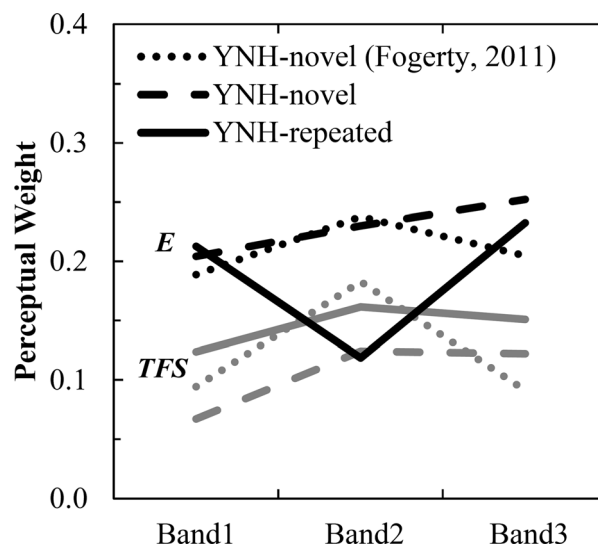

FIG. 7. Comparison of relative perceptual weights for YNH listeners. Black = E weights; Gray = TFS weights. Dotted lines display results for YNH listeners (of comparable age and hearing sensitivity) from Fogerty (2011a) who were tested on 600 novel sentences as a comparison to the novel and repeated sentence testing completed in Experiment 2 using 300 sentences each.

results for the four listener groups tested here. E cues still represent the primary cues, but in contrast to novel sentence testing, the most weight is placed on low- and high-frequency band E cues, with a relatively similar weight across the TFS channels.

These results suggest that when auditory feedback is provided by repeating the stimulus, listeners may "listen" to the repeated sentence differently by altering these relative weights. Thus, an initial encoding of the stimulus occurs by the "default," and potentially unstable, novel weighting pattern that is modified upon subsequent repetitions of the same stimulus. Individual differences between groups associated with age and hearing impairment are most apparent at that initial encoding, and subsequent repetitions result in similar perceptual weighting, and possibly similar perceptual encoding, of the stimulus.

D. Spectral and temporal contributions

Results from this study provide further demonstration that the temporal E is most susceptible to distortions in the speech signal. Such distortions could be due to noise, spectral shaping, cochlear filtering, or age-related declines in processing. In contrast, even though processing of the TFS has been found to be poorer in older adults (Grose and Mamo, 2010; Hopkins and Moore, 2011), particularly those with hearing loss (Lorenzi et al., 2006; Ardoint et al., 2010), the relative perceptual weights obtained for these temporal channels remained relatively stable across listener groups. Thus, perceptual use of the TFS appears relatively stable and not as susceptible to signal degradations. One additional example of TFS cues being robust against distortion is the case of temporal interruption of the speech signal. It appears that under such conditions, listeners rely more heavily upon the TFS of the stimulus (Nelson et al., 2003; Lorenzi et al., 2006), in part due to greater interference of temporal E cues from the external amplitude modulation imposed by the interrupting signal (Gilbert et al., 2007).

However, these results are at odds with the finding that older adults may be less able to process TFS cues (Grose and Mamo, 2010; Hopkins and Moore, 2011). Age did not appear to influence perceptual weighting of these cues. Either the older listeners in this study did not have declines in TFS processing, which is plausible given the results from Experiment 1B, or the concurrent availability of E cues was enough to facilitate sufficient extraction of cues from the degraded TFS signal. A direct investigation of TFS processing abilities and derived perceptual weighting is required.

However, even more significant is the need to investigate the relationship between speech recognition abilities and perceptual weighting. This study demonstrated that listeners may listen to speech differently and yet result in the same overall performance. As such, the overall percent correct measure may mask differences in performance between these listener groups. A more refined analysis of recognition errors and performance on stimulus subtypes may be necessary. However, such an analysis is outside the scope of the current study and must be left for future investigations. Likewise, the stability and adaptability of an individual's perceptual weights during stimulus training may also reveal a potential interaction between how listeners use different acoustic cues in the speech signal and their speech recognition abilities, particularly in noisy listening environments.

Finally, it is important to note that young hearing-impaired listeners were not included as one of the groups in this experiment. Therefore, a discussion of hearing impairment is derived from the older listeners in comparison to young normal-hearing controls who were noise-masked to obtain identical listening conditions.

VI. SUMMARY AND CONCLUSIONS

The results of the current study demonstrate the importance of E cues for young and older listeners, including those with hearing impairment. E cues were significantly weighted more than TFS cues for all listener groups during the presentation of novel sentences. However, the perceptual weighting of E cues appears to be highly susceptible to signal, and potentially cognitive, interactions. Previous work has demonstrated the influence of temporal interruption on E perceptual weights (Fogerty, 2011b). Here, results demonstrated that E weights are also influenced by age and spectral shaping of the stimulus (i.e., amplification as done by well fit hearing aids). As a caveat, the spectral shaping procedure employed a novel stimulus modification for both young spectrally-matched normal-hearing listeners and OHI listeners who had little to no experience with such a stimulus manipulation. Therefore, the shift to E cues in the low-frequency band observed for these groups with spectral shaping could potentially be a result of favoring "familiar" spectral cues and would not be observed for listeners with more experience with the modified spectral cues. Finally, limited experience with the sentences also influenced weighting of E cues, demonstrating a potential cognitive factor.

In contrast, TFS cues remained relatively stable across all conditions and listener groups. While these cues were weighted less by all listeners across all frequency bands for novel sentences, TFS weights appear to represent stable perceptual cues that are influenced to a limited degree by factors intrinsic to the signal or the listener. It remains to be determined how important the perceptual weighting of these cues is in determining speech recognition performance and if training of these cues would facilitate performance, particularly in adverse listening conditions.

ANSI, (1997). "Methods for the calculation of the speech intelligibility index," ANSI S3.5-1997 (American National Standards Institute, New York).

Apoux, F., and Bacon, S. P. (2004). "Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise," J. Acoust. Soc. Am. 116, 1671–1680.

Ardoint, M., Sheft, S., Fleuriot, P., Garnier, S., and Lorenzi, C. (2010). "Perception of temporal fine-structure cues in speech with minimal

envelope cues for listeners with mild-to-moderate hearing loss," Int. J. Audiol. **49**, 823–831.

Berg, B. G. (**1989**). "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am. **86**, 1743–1746.

Calandruccio, L., and Doherty, K. A. (**2008**). "Spectral weighting strategies for hearing-impaired listeners measured using a correlational method," J. Acoust. Soc. Am. **123**, 2367–2378.

Doherty, K. A., and Lutfi, R. A. (**1996**). "Spectral weights for overall level discrimination in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **99**, 1053–1058.

Doherty, K. A., and Lutfi, R. A. (**1999**). "Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **105**, 1831–1840.

Doherty, K. A., and Turner, C. W. (**1996**). "Use of a correlational method to estimate a listener's weighting function for speech," J. Acoust. Soc. Am. **100**, 3769–3773.

Fletcher, H., and Galt, R. (**1950**). "Perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.

Fogerty, D. (**2011a**). "Perceptual weighting of individual and concurrent cues for sentence intelligibility: Frequency, envelope, and fine structure," J. Acoust. Soc. Am. **129**, 977–988.

Fogerty, D. (**2011b**). "Perceptual weighting of the envelope and fine structure across frequency for sentence intelligibility: Effect of interruption at the syllabic-rate and periodic-rate of speech," J. Acoust. Soc. Am. **130**, 489–500.

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.

French, N. E., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (**1990**). "DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM," National Institute of Standards and Technology, NTIS Order No. PB91-505065.

Gilbert, G., Bergeras, I., Voillery, D., and Lorenzi, C. (**2007**). "Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues," J. Acoust. Soc. Am. **122**, 1336–1339.

Greenberg, S., Arai, T., and Silipo, R. (**1998**). "Speech intelligibility derived from exceedingly sparse spectral information," in *Proceedings of the International Conference on Spoken Language Processing*, pp. 2803–2806.

Grose, J. H., and Mamo, S. K. (**2010**). "Processing of temporal fine structure as a function of age," Ear Hear. **21**, 755–760.

Hedrick, M. S., and Younger, M. S. (**2007**). "Perceptual weighting of stop consonant cues by normal and impaired listeners in reverberation versus noise," J. Speech Lang. Hear. Res. **50**, 254–269.

Hopkins, K., and Moore, B. C. J. (**2010**). "The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **127**, 1595–1608.

Hopkins, K., and Moore, B. C. J. (**2011**). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," J. Acoust. Soc. Am. **130**, 334–349.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

IEEE. (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

Leibold, L., Tan, H., and Jesteadt, W. (**2006**). "Spectral weights for level discrimination in quiet and in noise," Association for Research in Otolaryngology, Mid-Winter Meeting, Baltimore, MD.

Liberman, M. C. (**1982**). "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," J. Acoust. Soc. Am. **72**, 1441–1449.

Loizou, P. C. (**2007**). *Speech Enhancement: Theory and Practice* (CRC Press, Taylor and Francis, Boca Raton, FL), pp. 1–608.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Lutfi, R. A. (**1995**). "Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks," J. Acoust. Soc. Am. **97**, 1333–1334.

Mehr, M. A., Turner, C. W., and Parkinson, A. (**2001**). "Channel weights for speech recognition in cochlear implant users," J. Acoust. Soc. Am. **109**, 359–366.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation pattern," J. Acoust. Soc. Am. **74**, 750–753.

Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.

Pittman, A. L., and Stelmachowicz, P. G. (**2000**). "Perception of voiceless fricatives by normal- hearing and hearing-impaired children and adults," J. Speech Lang. Hear. Res. **43**, 1389–1401.

Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.

Richards, V. M., and Zhu, S. (**1994**). "Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients," J. Acoust. Soc. Am. **95**, 423–434.

Seewald, R. C., Ramji, K. V., Sinclair, S. T., Moodie, K. S., and Jamieson, D. G. (**1993**). *Computer-Assisted Implementation of the Desired Sensation Level Method for Electroacoustic Selection and Fitting in Children: Version 3.1, User's Manual* (The University of Western Ontario, London).

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Souza, P. E., and Boike, K. T. (**2006**). "Combining temporal-envelope cues across channels: Effects of age and hearing loss," J. Speech Lang. Hear. Res. **49**, 138–149.

Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech- transmission quality," J. Acoust. Soc. Am. **67**, 318–326.

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

Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.