

Research Article

Contributions to Speech-Cue Weighting in Older Adults With Impaired Hearing

Pamela Souza,^a Frederick Gallun,^b and Richard Wright^c

Purpose: In a previous paper (Souza, Wright, Blackburn, Tatman, & Gallun, 2015), we explored the extent to which individuals with sensorineural hearing loss used different cues for speech identification when multiple cues were available. Specifically, some listeners placed the greatest weight on spectral cues (spectral shape and/or formant transition), whereas others relied on the temporal envelope. In the current study, we aimed to determine whether listeners who relied on temporal envelope did so because they were unable to discriminate the formant information at a level sufficient to use it for identification and the extent to which a brief discrimination test could predict cue weighting patterns. **Method:** Participants were 30 older adults with bilateral sensorineural hearing loss. The first task was to label synthetic speech tokens based on the combined percept of temporal envelope rise time and formant transitions. An individual profile was derived from linear discriminant analysis of the identification responses. The second task was to discriminate

differences in either temporal envelope rise time or formant transitions. The third task was to discriminate spectrotemporal modulation in a nonspeech stimulus.

Results: All listeners were able to discriminate temporal envelope rise time at levels sufficient for the identification task. There was wide variability in the ability to discriminate formant transitions, and that ability predicted approximately one third of the variance in the identification task. There was no relationship between performance in the identification task and either amount of hearing loss or ability to discriminate nonspeech spectrotemporal modulation.

Conclusions: The data suggest that listeners who rely to a greater extent on temporal cues lack the ability to discriminate fine-grained spectral information. The fact that the amount of hearing loss was not associated with the cue profile underscores the need to characterize individual abilities in a more nuanced way than can be captured by the pure-tone audiogram.

Speech information is transmitted by a complex set of concomitant acoustic cues (Stevens, 1997). Shaping of the vocal tract via movement of the articulators defines overall spectral shape and formant transitions to convey information about vowel identity, consonant place, and consonant manner. The time-varying constriction of airflow by the lips and tongue modifies signal amplitude to convey information about consonant manner, voicing, and place. Acoustic redundancy supports speech recognition even when some cues are degraded or removed by difficult listening scenarios, such as noise or reverberation.

However, cue redundancy can be offset by limitations to the listener's ability. One example of this is a listener with sensorineural hearing loss who has a reduced ability to receive and interpret acoustic speech cues. While the obvious consequence of hearing loss is inadequate audibility, many listeners with sensorineural loss also have difficulty discriminating suprathreshold information. Indeed, complaints about speech discrimination have persisted in this population despite decades of well-fit, high-fidelity hearing aids that largely correct for audibility deficits (Kochkin, 2007; McCormack & Fortnum, 2013).

It has been assumed that, when the output of sensory processing for speech is distorted, one cause is cochlear damage that broadens the precise tuning present in a normal cochlea. This assumption is supported by group data that hearing loss results in loss of spectral resolution (Davies-Venn, Nelson, & Souza, 2015; Glasberg & Moore, 1986; Moore, Vickers, Plack, & Oxenham, 1999; Souza & Hoover, 2018). However, other types of damage may also occur and may have different perceptual consequences. Noise damage results in loss of spiral ganglion fibers, even in cases of intact outer hair cells (Kujawa & Liberman, 2009). Inner

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

^bRehabilitation Research and Development National Center for Rehabilitative Auditory Research, VA Portland Health Care System and Oregon Health and Sciences University

^cDepartment of Linguistics, University of Washington, Seattle

Correspondence to Pamela Souza: p-souza@northwestern.edu

Editor-in-Chief: Bharath Chandrasekaran

Editor: Chao-Yang Lee

Received December 21, 2019

Revision received June 21, 2019

Accepted October 2, 2019

https://doi.org/10.1044/2019_JSLHR-19-00176

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

hair cells may be damaged in focal lesions or in diffuse patterns (Dubno, Eckert, Lee, Matthews, & Schmiedt, 2013; Stebbins, Hawkins, Johnson, & Moody, 1979). Aging may impair neural synchrony, affecting use of temporal fine structure (Pichora-Fuller, Schneider, Macdonald, Pass, & Brown, 2007). In addition, some listeners who have a history of traumatic brain injury may be left with impaired temporal perception (Hoover, Souza, & Gallun, 2017; Turgeon, Champoux, Lepore, Leclerc, & Ellemberg, 2011). Each type of damage will have unique effects on fidelity of suprathreshold speech information, and the aggregate damage will vary by individual. For those reasons, it is not possible to accurately predict individual suprathreshold resolution from the pure-tone audiogram.

To explore individual use of suprathreshold speech information, we devised a test to measure listeners' use of speech cues when both temporal and spectral cues were available (Souza, Wright, Gallun, & Reinhart, 2018; Souza, Wright, Blackburn, Tatman, & Gallun, 2015). Test materials consisted of synthetic speechlike syllables in which sets of speech cues (either temporal envelope rise time and spectral shape or temporal envelope rise time and formant transitions) were manipulated to have specific parameters. We hypothesized that a listener who used spectral information would systematically identify different sounds as the spectral dimension was manipulated. In contrast, a listener who was insensitive to spectral differences would maintain the same percept as the spectral dimension was manipulated. A similar relationship was hypothesized to occur for temporal envelope. Analysis of the relative sensitivity to spectral cues versus temporal envelope across the entire matrix of tokens (referred to here as the "cue profile") captured distinctive aspects of the listener's use of suprathreshold speech cues.

Because the cochlear and neural damage patterns caused by sensorineural hearing loss are expected to vary across individuals, we hypothesized that the ability to use specific cues for speech recognition would also vary across individuals and would not be strongly predicted by pure-tone thresholds. These hypotheses were supported by our data. When spectral information was relatively static and broad in frequency (i.e., conveyed by spectral shape), 90% of the listeners used the spectral information to identify the phoneme. However, when spectral information was dynamic and fine-grained (i.e., conveyed by formant transitions), only 50% of the listeners reliably used the spectral information to identify the phoneme. The remaining listeners categorized the phonemes primarily on the basis of temporal envelope rise time, regardless of what formant information was present, and were therefore considered to rely to a greater extent on temporal envelope. We and others (e.g., Boothroyd, Springer, Smith, & Schulman, 1988) have assumed that such listeners were unable to discriminate spectral detail. That is, a greater reliance on temporal envelope was obligatory rather than a preferential or habitual listening strategy.

Such an obligatory listening pattern might direct treatment options. Several authors have suggested that listeners who rely on temporal envelope should receive hearing aids

that preserve fidelity of temporal envelope information (Angelo & Behrens, 2014; Sabin & Souza, 2013; Weile, Behrens, & Wagener, 2011). That approach might dictate longer compression time constants, aggressive suppression of temporal-distorting reverberation, or envelope enhancement. However, direction of treatment must also consider what factors lead to an apparent temporal cue reliance. For example, if reliance on temporal cues represents a strategy rather than a deficit, it may be possible to train listeners to use a more informative cue (Lowenstein & Nittrouer, 2015; Moberly et al., 2014; Toscano & McMurray, 2010) or enhance the unused cue to make it more salient (Leek & Summers, 1996).

The cue profile task used in our earlier studies required approximately an hour of testing plus statistical expertise to model the cue profile using linear discriminant analysis (LDA). An additional interest of our group was to determine whether a brief task in which listeners were asked to discriminate only the spectral (or only the temporal) aspects of the stimuli could reliably predict a listener's cue profile. Such a test would take less time and be easier to interpret than the full identification protocol. In addition, a test that depended on perceiving differences between stimuli (rather than on labeling them) could be administered without extensive training and to listeners with different language backgrounds.

To address these goals, we created adaptive discrimination tasks that were derived from the same synthetic speechlike syllables as used in the identification task. The discrimination task measured whether or not the listener could perceive the difference between two adjacent stimuli. If no difference can be perceived, then it is not surprising that the same label is applied to both. This level of precision was sufficient for asking the question of whether or not identification was limited by discrimination ability. (Note, however, that our adaptive discrimination tasks were not discrimination in the sense of measuring the smallest possible acoustic difference that the listener could perceive.)

We also measured ability to discriminate spectrotemporal modulation in a nonspeech signal. Sensitivity to time-varying spectral "ripple" has been shown to be strongly related to sentence recognition in noise (e.g., Bernstein et al., 2016, 2013; Mehraei, Gallun, Leek, & Bernstein, 2014; Miller, Bernstein, Zhang, Wu, Bentler, & Tremblay, 2018). Because discriminating differences in time-varying spectral ripple has some similarity to discriminating differences in time-varying formant information, we explored whether a version of the spectrotemporal modulation task already adapted for clinical use (Gallun et al., 2018) could predict a listener's cue profile.

Method

Participants

Thirty older adults (12 women, 18 men) with bilateral hearing loss participated in the study. Mean listener age was 74 years (range: 61–90 years). All listeners were native speakers

of American English who had symmetrical sensorineural loss with nonsignificant (≤ 10 dB) air–bone gaps. Audiograms are shown in Figure 1. The mean difference in pure-tone average between right and left ears was 3 dB (range: 0–13 dB). Twelve listeners wore hearing aids on a regular basis (11 bilaterally aided and one unilaterally aided in his poorer ear). Those participants' hearing aids were 3 years old on average (range: 1–6 years). All listeners passed a cognitive screening (Nasreddine et al., 2005).

Materials

The identification and the discrimination tests used the same set of synthetic syllables, created by controlling temporal envelope rise time for the temporal dimension in combination with formant transitions for the spectrotemporal dimension. Parameter values for those stimuli were based on Klatt (1980) and implemented via the Synthworks implementation of the Klatt parametric synthesizer (v. 1112) on a Macbook Pro (Mac OSX 10.7.4). The formant frequencies were varied in five steps, ranging from formants typical of alveolar consonants (/l/ or /d/) to formants typical of labial consonants (/w/ or /b/). Amplitude rise time and formant transition time were also varied in five steps, ranging from envelopes typical of approximant consonants (/l/ or /w/) to envelopes typical of stop consonants (/b/ or /d/; see Table 1). In all tokens, the formant frequencies and bandwidths of the vowel portion of the syllable represented the vowel /a/ spoken by a male talker.

The combination of Spectral (e.g., formant transitions) \times Temporal (e.g., envelope rise time) parameters created a set of 25 unique tokens (see Figure 2a). The four most extreme conditions (referred to here as “end points” see Figure 2b) were most readily associated with specific syllables. For example, the tokens labeled “da” and “la” had the same formant onset frequencies and transitions but different temporal envelope rise times (10 ms for “da” and 60 ms for

“la”). Although there was no a priori correct label for the three stimuli represented by the cells between “da” and “la,” we expected that listeners who were able to perceive differences in temporal envelope would shift from perception of “da” to perception of “la” as envelope rise time was increased. Similarly, the tokens labeled “la” and “wa” had the same temporal envelope rise time but different formant transitions (e.g., F2 onset 1500 Hz for “la” and 850 Hz for “wa”). Although there was no a priori correct label for the three stimuli represented by the cells between “la” and “wa,” we expected listeners who were able to perceive differences in formant transitions to shift from perception of “la” to perception of “wa” as F2 (and the corresponding F1, F3, and F4 values; see Table 1) changed. Similar effects as determined by the listener's sensitivity to temporal envelope and to formant transitions were expected for stimuli represented by the remaining cells in Figure 2a.

The stimuli were converted from digital to analog (TDT RX6), level adjusted using a programmable attenuator (TDT PA5), and presented through a headphone buffer (TDT HB7) to an insert earphone (ER-2). One ear of each participant was selected for testing. This was generally the right ear, but in some participants, the left ear was selected as the test ear if adequate audibility could not be achieved in the right ear. To compensate for loss of audibility due to threshold elevation, a flat individual gain was used to ensure audible bandwidth through 5 kHz (corresponding to the bandwidth of the test stimuli). Individual frequency shaping was deliberately not used to avoid confounding the desired spectrotemporal manipulations.

Procedure

All participants completed an informed consent process and were reimbursed at an hourly rate for their time. Testing was conducted over two visits of approximately 1.5–2 hr each, including frequent rest breaks. Auditory testing was conducted in a double-walled, sound-treated booth.

Stimulus Identification

Participants completed training to associate the synthetic speech sounds with labels. The training presented only the four “end point” stimuli in which the most extreme temporal and spectral characteristics were combined. First, to familiarize them with the test stimuli, the listener heard 40 (10 repetitions \times 4 end points) randomly ordered trials in which an end point was played while the correct alternative automatically highlighted green. Next, the listener heard 40 randomly ordered trials in which an end point was played and the task was to identify that token from the set of four possible alternatives. Participants used a computer mouse to select the stimulus they heard from a display of four alternatives, labeled orthogonally as “dah,” “lah,” “bah,” and “wah.” The four end points were equally represented, and visual correct answer feedback was provided. If the participant scored at least 80% correct for each end

Figure 1. Left and right ear audiograms for the tested group. Thin lines show individual audiograms; thick lines show the mean audiogram for that ear.

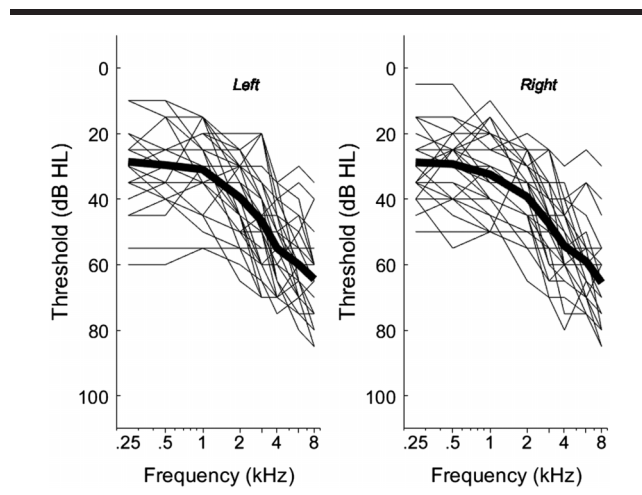


Table 1. Spectral and temporal parameters for test stimuli.

Phoneme	/d/	/b/	/l/	/w/
Temporal	Stop	Stop	Approximant	Approximant
Spectral	Alveolar	Labial	Alveolar	Labial
F1 (Hz)	250	250	250	250
F2 onset (Hz)	1500	850	1500	850
F3 onset (Hz)	2400	2150	2400	2150
F4 (Hz)	3250	3250	3250	3250
Temporal envelope rise time (ms)	10	10	60	60
Formant transition (ms)	40	40	110	110

point, training was complete and the testing phase began. If not, another block of 40 randomly selected end point trials was presented. The training blocks repeated until the participant reached 80% correct for each end point or had completed four blocks of 40 end point trials, whichever occurred first.

After training was complete, the participant completed 375 trials, which included 15 trials of each of the 25 tokens, presented in random order without correct answer feedback. Rest breaks were given after 125 and 250 trials.

Both independent (spectral and temporal parameter steps) and dependent measures (stimulus identification) were

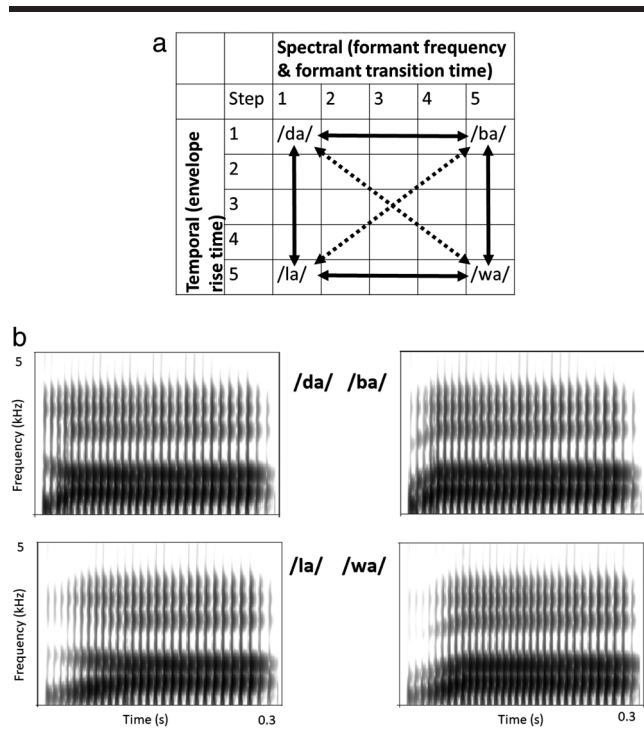
discrete steps along a continuum. Accordingly, data were analyzed via an LDA (Lachenbruch & Goldstein, 1979). The LDA output is an equation whose coefficients quantify how much each of the predictors contributed to the classification (Hair, Black, Babin, & Anderson, 2010). The standardized coefficients, expressed in absolute value, allowed us to interpret the relative importance of the formant transitions and temporal envelope rise time to the listener's responses on the same scale. The LDA also provided the proportion of between-groups variance accounted for by the given equation (i.e., the goodness of fit).

Stimulus Discrimination

Equipment, presentation level, and test ear for each participant were as described for the identification task. As noted in the introduction, the purpose of this test was to measure the listener's ability to make distinctions along each of the dimensions included in the design of the stimuli and identified by the LDA. By varying only formant transition or only temporal envelope rise time and requiring a same/different judgment, the same stimuli could be tested without requiring the listener to assign a label to a perceived multidimensional stimulus. In that way, we wished to understand what stimulus distinctions the listener could (or could not) discriminate and use when performing the stimulus identification task.

A four-interval two-cue, two-alternative forced-choice design was used in order to ensure that the discrimination was limited by sensory information rather than memory or attention. Four buttons were displayed on the computer monitor, corresponding to the four intervals. The buttons were highlighted (changed color) in sequence as the four sounds were played, and the observer knew that, when the first and fourth buttons were highlighted, the "standard" sound would be played. During either the second or third interval, a "test" sound was played, with a standard sound played in the other interval. A trial was defined as the presentation of the four intervals, each interval lasting 300 ms, with 800 ms between intervals. Thus, during each trial, the listener heard four 300-ms synthetic speech tokens: three identical standard tokens and one "oddball" test token. The task was to click on the button that was highlighted when the test sound was played. This paradigm reduces attention and memory impacts on performance because it ensures that the test stimulus is always presented with a standard

Figure 2. (a) Schematic representation of stimuli for the identification and discrimination tasks. The combination of 5 spectral (e.g., formant values) \times 5 temporal (e.g., amplitude rise time) parameters created a set of 25 unique tokens, represented by the 5×5 grid. The relationship between signals along the solid and dashed arrows are explained in the text. (b) Spectrograms for the four "end point" stimuli represented by the corner cells in (a).



stimulus preceding it and following it (i.e., the task can be performed by comparing either forward or backward in time).

The familiarization phase consisted of two blocks of trials. A different standard was used in each block. Standards were one of the two tokens that were most dissimilar from each other on both temporal and spectral dimensions (e.g., the cells marked “ba” and “la” in Figure 2a). These tokens were chosen as the standards for the familiarization in order to confirm that the participant understood the instructions and could detect differences in at least one dimension (temporal envelope rise time and/or formant transition). The initial trial of each block included one of these stimuli as the standard and the other as the “different” comparison token. Threshold was estimated based on a two-down/one-up adaptive track. That is, after two correct responses, the comparison was moved one step closer to the standard in similarity for both temporal and spectral dimensions, while after every incorrect response the comparison was moved one step away from the standard (along the dashed line in Figure 2a). Whenever the step size had been getting smaller and then got larger or had been getting larger and then got smaller, the program marked the stimulus at which the change occurred as a “reversal.” The maximum distance was four steps, and the minimum distance was one step, which was also the smallest distinction present in the cue profile test. The track continued until 10 reversals were marked and the step sizes corresponding to the last six reversals were averaged to produce a threshold distance estimate. As it was anticipated that some participants would be able to distinguish all of the stimuli in the matrix shown in Figure 2a, blocks were terminated when four correct responses in a row were obtained at the smallest step size. In that case, threshold was estimated to be a step size of 1. Each participant completed two familiarization blocks, one in which the “ba” stimulus was the standard and one in which the “la” stimulus was the standard, in random order.

The test phase consisted of 24 blocks. In 12 of the blocks, the participant was asked to discriminate stimuli that differed only in envelope rise time (represented by the vertical solid arrows in Figure 2a). As an example, a given trial for stimuli that varied in envelope rise time might use end point “da” (Column 1, Row 1 in the matrix in Figure 2a) as the standard, initially compared with “la” (Column 1, Row 5). If correctly discriminated, the test would adapt to compare Row 1 with Row 4, then Row 1 with Row 3, and then Row 1 with Row 2. The test would end with the listener having successfully discriminated to the smallest temporal envelope step size when each temporal envelope was combined with the same formant transition. In the other 12 blocks, the participant was asked to discriminate stimuli that differed only in formant transition (represented by the horizontal solid arrows in Figure 2a) for the same envelope rise time. Because stimuli were always discriminated relative to a standard, only the edges of the matrix were tested. Each end point was equally represented as the standard, and the 24 blocks were presented in random order. The final

spectral and temporal thresholds were calculated as the mean of the smallest step size the listener was able to discriminate across 12 blocks (4 end points \times 3 repetitions) for each dimension.

Sensitivity to Nonspeech Spectrotemporal Modulation

A variation of the methods employed by Bernstein et al. (2013) was used to measure the lowest modulation depth at which listeners could detect spectrotemporal modulation with a spectral modulation rate of 2 cycles/octave in the presence of temporal modulation at a rate of 4 Hz. This combination of spectral and temporal modulations results in a stimulus in which the spectral peaks move either up or down in frequency. As Bernstein et al. found that the direction of motion did not affect sensitivity, direction was randomized across stimuli. Stimuli were presented to the test ear (same ear as for stimulus identification and discrimination) via a custom iPad app (Gallun et al., 2018) coupled to Sennheiser HD 280Pro earphones. Listeners first performed an audibility task in order to set the levels for the modulation detection task. Audibility was measured using the same type of four-interval two-cue, two-alternative forced-choice paradigm as was used in the speech token discrimination task. Four buttons were displayed in a graphical user interface on the iPad screen corresponding to the four intervals, and each was sequentially highlighted for 500 ms, with 250 ms between the intervals during which none of the buttons were highlighted. While the first and fourth buttons were highlighted, no sound was presented. This was the standard stimulus. While either the second or third button was highlighted, a noise burst of 500 ms in duration was presented, and the task of the listener was to identify the interval during which sound was presented by pressing the button that had been highlighted when they heard the noise. The level of the noise burst was adaptively changed in order to determine the audibility threshold. Noise bursts were broadband, flat amplitude noise between 400 and 8000 Hz and were thus spectrotemporally identical to the unmodulated stimuli to be used as standards in the modulation detection task (described below). Adaptive tracks started at 50 dB SPL, and the level was reduced by 5 dB after every two correct responses and raised by 10 dB SPL after every incorrect response until the changes in level had reversed twice, at which point the steps up became 5 dB, the steps down became 2 dB, and six more reversals were obtained. Detection threshold was defined as the average of the levels at which the final six reversals occurred. Once the just-audible level had been established, the levels for the stimuli to be used in the modulation detection task could be set. To ensure equal audibility across listeners, levels were set 30 dB above the detection threshold. For our cohort, the mean detection threshold was 45.0 dB SPL (range: 23.8–65.5 dB SPL), and the modulated stimuli were presented at a mean presentation level of 75.0 dB SPL (range: 53.8–85.5 dB SPL). To remove cues associated with differences in loudness between modulated and unmodulated stimuli, the level of each stimulus was then roved within a range of ± 3 dB on each presentation interval.

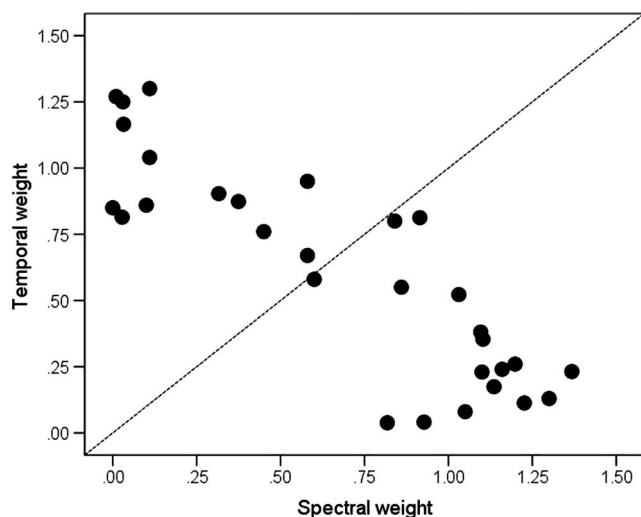
The spectrotemporal modulation was implemented with sinusoidal modulation on a log scale (dB) in both temporal and spectral domains. As with the audibility task and the token discrimination task, a four-interval two-cue, two-alternative forced-choice paradigm was used. The standard stimulus was flat spectrum noise of the type presented in the detection task, and the task of the listener was to identify whether the second or third interval contained noise that had been spectrotemporally modulated. Modulation depth was adaptively varied again using a two-down/one-up tracking procedure to provide an estimate of modulation threshold. Two reversals were obtained, with increases in modulation depth of 1 dB after every incorrect response and decreases of 0.5 dB after every two correct responses. Then, six more reversals were obtained with increases of 0.2 dB and decreases of 0.1 dB. The threshold was defined as the average of the modulation depths at which the final six reversals were obtained, where modulation depth refers to the decibel difference between the peaks and valleys of the modulated stimulus. The final threshold was the mean across two blocks.

Results

Stimulus Identification

Figure 3 shows the points in the LDA space, plotted as the absolute value of the standardized coefficient or “weight” for the temporal dimension as a function of the absolute value of the standardized coefficient for the spectral dimension. Spectral and temporal weights were strongly

Figure 3. Coefficient weights derived from the linear discriminant analysis. Each point represents a listener. Points falling below the diagonal (weighting angle $< 45^\circ$) show a greater influence of formant information to syllable identification. Points falling above the diagonal (weighting angle $> 45^\circ$) show a greater influence of temporal envelope information to syllable identification. Points falling along or close to the diagonal suggest approximately equal influence of formant and temporal envelope information to syllable identification.



related to each other, such that listeners who relied on spectral information were minimally influenced by temporal information and vice versa (Pearson $r = -.86$, $p < .001$). Each point location was specified as a “weighting angle,”¹ using the following equation:

$$\text{Weighting angle} = \tan^{-1}(\text{temporal weight}/\text{spectral weight}). \quad (1)$$

Each angle describes a location in the space, with 0° representing that syllable identification is based on spectral information, 90° representing that syllable identification is based on temporal information, and 45° representing that temporal and spectral information contribute equally to syllable identification. Individual weights ranged from 1.6° to 89.6° ($M = 43.8^\circ$). This range is consistent with previous work (Souza et al., 2018; range = 2.1° – 90.0° , $M = 53.9^\circ$).

Stimulus Discrimination

Individual spectral discrimination thresholds ranged from 1 to 3.9 steps ($M = 1.9$ steps). Individual temporal discrimination thresholds ranged from 1 to 1.9 steps ($M = 1.2$ steps). The relationship between spectral discrimination thresholds and the cue profile is plotted in Figure 4. In recognition that the score for the discrimination task is the best (i.e., smallest) stimulus difference that can be perceived rather than a fine-grained discrimination threshold, discrimination is categorized here according to step size, with smaller step sizes representing better performance. For the cue profile, weighting angles closer to 0° represent that the listener’s syllable identification is more strongly influenced by formant transition, and weighting angles closer to 90° indicate that the listener’s syllable identification is more strongly influenced by temporal envelope rise time. The data suggest that better discrimination (i.e., the ability to discriminate a smaller change in formant information) is associated with stronger use of that cue in the identification task. The implication is that reliance on temporal envelope is associated with an inability to discriminate the spectral (formant) information.

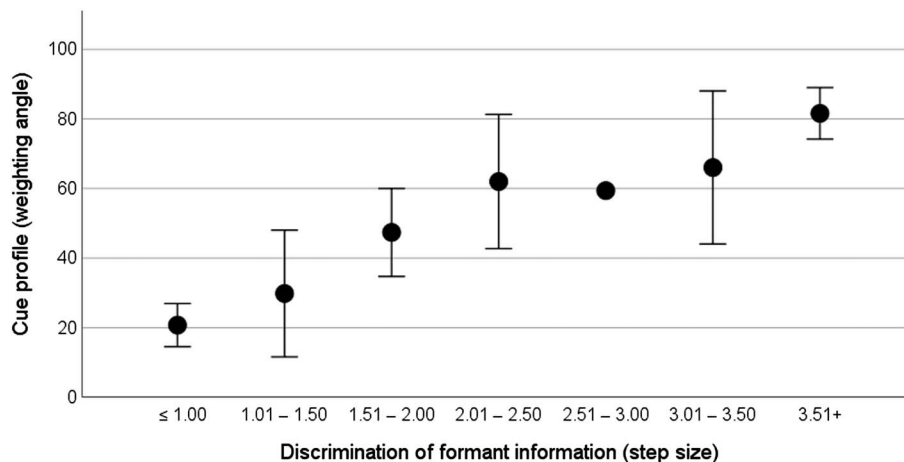
The relationship between temporal discrimination thresholds and the cue profile is plotted in Figure 5, again with discrimination categorized according to step size, with smaller step sizes representing better performance. No individual had a temporal discrimination threshold poorer than two steps, and there was not a strong relationship between the temporal discrimination threshold and the cue profile.

Sensitivity to Spectrotemporal Modulation

The mean modulation threshold was 4.0 dB (range: 1.1–10.1 dB). Listeners with poorer hearing (expressed as

¹The magnitude of the vector associated with the weighting angle describes an additional dimension, which is not considered in this analysis.

Figure 4. Relationship between spectral discrimination (categorized by step size with smaller step sizes representing better performance) and weighting angle. Circles show the mean weighting angle, and error bars show ± 1 SE about the mean. Weighting angles closer to 0° represent that the listener's syllable identification is more strongly influenced by formant trajectory, and weighting angles closer to 90° indicate that the listener's syllable identification is more strongly influenced by temporal envelope.



pure-tone average) had poorer detection thresholds ($r = .73$, $p < .005$) and reduced sensitivity to spectrotemporal modulation ($r = .71$, $p < .005$). Modulation thresholds were stable (mean differences < 1 dB) with retest on a different day in a subset ($n = 7$) of participants.

Relationships Among Variables

As a first step in exploring relationships among the variables, Table 2 shows Pearson correlations² among cue weighting angle, cue discrimination threshold, sensitivity to spectrotemporal modulation, and amount of hearing loss. Bold font indicates relationships that were significant at $p < .001$. The correlation values suggested that individual ability to discriminate differences in formant transition was the strongest determinant of whether spectral or temporal cues were used for speech identification. The ability to utilize formant information was related to sensitivity, to spectrotemporal modulation, and to the amount of hearing loss.

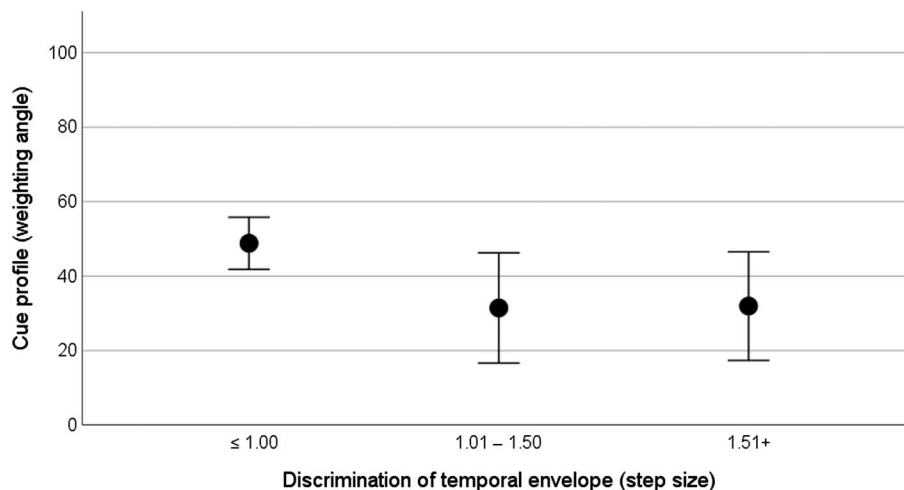
Because spectral discrimination was significantly associated with the outcome in univariable models (i.e., correlations), we used a multivariable model to assess whether spectral discrimination remained significantly associated with the outcome after adjustment for other predictors. A hierarchical regression model was built with predictors entered in the following order and based on these rationales. (a) Pure-tone average: If the cue profile were adopted for clinical testing, pure-tone thresholds would be the most likely predictor to be available for all listeners. (b) Discrimination threshold for formant transition: Consistent with our hypotheses and supported by high correlations in the univariable models, poor formant discrimination was expected to be related to weighting angle. (c) Discrimination

threshold for temporal envelope rise time: This was included in the model to account for any performance restriction due to poor temporal envelope perception, although we did not expect that to be a significant predictor. (d) Spectrotemporal modulation: Unlike (b) and (c), this discrimination threshold was measured to the smallest possible modulation difference. It was included to account for any additional variance but was entered last based on the univariable models, which showed no relationship with weighting angle. The dependent variable was the weighting angle (which expressed the relative importance of spectral and temporal cues to the listener's ability to identify speech sounds). Residual and scatter plots indicated that the assumptions of normality and linearity were satisfied and collinearity within the model was acceptable (variance inflation factors < 2.0 , tolerances $> .5$)³. Results are shown in Table 3. The initial model (Step 1), based on pure-tone average, failed to significantly explain variance, $F(1, 29) = 0.50$, $p = .487$, consistent with earlier findings that amount of hearing loss did not predict the cue profile (Souza et al., 2018). When spectral discrimination was entered (Step 2), the model was significant, $F(2, 29) = 6.62$, $p = .005$, and the change in variance accounted for was significant. The addition of temporal discrimination (Step 3) accounted for about 6% more variance, $F(3, 29) = 5.39$, $p = .005$. Adding spectrotemporal modulation thresholds as a predictor (Step 4) did not account for significant additional variance, although the overall model remained significant, $F(4, 29) = 4.01$, $p = .012$.

³For a valid analysis, tolerance ($1 - R^2$, where R^2 is calculated by regressing each independent variable on the other independent variables) should be .10 or higher, and the variance inflation factor (the increase in standard errors compared to a situation where the predictors are not correlated) should be less than 10 (Hair et al., 2010).

²These were uncorrected for multiple comparisons.

Figure 5. Relationship between temporal discrimination (categorized by step size with smaller step sizes representing better performance) and weighting angle. Circles show the mean weighting angle, and error bars show ± 1 SE about the mean. Weighting angles closer to 0° represent that the listener's syllable identification is more strongly influenced by formant trajectory, and weighting angles closer to 90° indicate that the listener's syllable identification is more strongly influenced by temporal envelope.



The final model accounted for 39% of the variance, with almost all of it attributed to spectral discrimination.

Discussion

The data reported here shed light on a persistent issue: Why do listeners with similar audiograms have such varied speech recognition abilities? By focusing on context-free speech materials with controlled acoustic cues, the task eliminated possible confounds such as linguistic complexity and contextual information. Instead, the task required the listener to discriminate specific speech cues and apply those cues to decide which sound was heard. The cues considered here—temporal envelope rise time and formant transition—were selected as key features of speech perception in everyday listening that would be expected to be extended to other recognition scenarios.

Two previous articles in this line of work (Souza et al., 2015, 2018) indicated that individuals with sensorineural hearing loss were influenced by different cues to identify speechlike sounds. Some listeners used formant cues to speech identification, whereas others relied on overall amplitude. The genesis of those patterns was unclear, particularly whether listeners who depended to a relatively greater extent on temporal cues to speech did so because poor resolution limited access to spectral cues. Other authors have addressed that question with relatively static spectral cues, such as distinguishing time-invariant vowel formants. By creating synthetic stimuli with dynamic formant information, we focused on the dynamic spectral information, which is more characteristic of that which would occur in everyday speech and which provides a higher level of information (Nittrouer, Lowenstein, Wucinich, & Tarr, 2014; Viswanathan, Magnuson, & Fowler, 2014).

Table 2. Pearson correlation among variables.

Variable	Identification (weighting angle)	Discrimination (spectral)	Discrimination (temporal)	Spectrotemporal modulation threshold	Pure-tone average
Identification (weighting angle)	<i>r</i> 1	.56	-.20	.30	.13
	<i>p</i>	.001	.291	.109	.487
Discrimination (spectral)	<i>r</i>	1	.09	.62	.43
	<i>p</i>		.650	< .001	.017
Discrimination (temporal)	<i>r</i>		1	.29	.16
	<i>p</i>			.115	.388
Spectrotemporal modulation threshold	<i>r</i>			1	.71
	<i>p</i>				< .001
Pure-tone average	<i>r</i>				1
	<i>p</i>				

Note. Relationships that are significant at $p < .001$ are indicated in bold.

Table 3. Summary of regression analysis for variables predicting syllable identification.

Step no.	Variable	β	t	R	R^2	ΔR^2	F	p	pr
Step 1	PTA	.13	0.71	.13	.02	.02	0.50	.487	.13
Step 2	PTA	-.13	-0.77	.57	.33	.31	12.53	.445	-.15
	Spectral discrimination	.62	3.54					.001	.56
Step 3	PTA	-.10	-0.57	.62	.38	.05	2.30	.141	-.11
	Spectral discrimination	.63	3.65					.001	.57
	Temporal discrimination	-.24	-1.52					.141	-.29
Step 4	PTA	-.17	-0.78	.63	.39	.01	0.29	.592	-.15
	Spectral discrimination	.57	2.82					.009	.49
	Temporal discrimination	-.26	-1.59					.124	-.30
	Spectrotemporal modulation	.15	0.54					.592	.11

Note. F and p values in this table refer to the effect of adding additional variables. F and p values for the overall models at each step are given in the text. Relationships that are significant at $p < .001$ are indicated in bold. PTA = pure-tone average.

Format Discrimination and the Cue Profile

The present results indicated that about one third of the variance in identifying phonemes distinguished by formant and amplitude information was accounted for by the listener's ability to discriminate formant information. These findings are consistent with the work of Shrivastav, Humes, and Kewley-Port (2006), who found that discrimination of spectral shape in manipulated phonemes explained about 34% of the variance in identification of naturally produced syllables. As was the case here, Shrivastav et al. also presented their speech stimuli under conditions of good stimulus audibility. Taken together, these two articles suggest that the ability to adequately resolve static and dynamic spectral information in audible speech signals is an essential precursor to using that information for speech identification, that this ability varies among individuals but does not depend on amount of hearing loss, and that individuals who cannot resolve spectral cues to speech will have poorer overall identification, regardless of their degree of hearing loss.

There have been many informative laboratory tests that could be used to guide treatment decisions but were rarely or never used in clinical practice because they took too much time. The relationship between discrimination of spectral cues (which takes less than 5 min to measure) and the more extensive cue profile may facilitate translation of our measures to a clinical environment. The discrimination test also allows for straightforward interpretation, consisting of a single "threshold" value that is associated with use of that cue in broader listening tasks.

The finding that the amount of hearing loss was not associated with the ability to resolve formant cues is interesting, because it has been generally thought that listeners with sensorineural loss (and presumed damage to the outer hair cells) will experience loss of spectral resolution and that poorer pure-tone thresholds will be associated with greater spectral degradation for speech sounds (e.g., Larsby & Arlinger, 1999; Leek, Dorman, & Summerfield, 1987; Molis

& Leek, 2011). In the group tested here, some listeners were able to discriminate small formant differences even when they had more hearing loss. Those listeners usually—but not always—relied upon formant transition information to identify sounds, although some listeners who apparently had the ability to discriminate formant transition information did not use it for speech identification.

Other Contributions to the Cue Profile

For the listeners tested here, discrimination of temporal envelope rise time accounted for a small amount of variance in phoneme identification that did not improve model fit. In contrast to the wide range of individual ability to discriminate formant transitions, nearly all listeners were good discriminators of temporal envelope for the contrasts tested here. Those contrasts were relatively coarse (rise time differences on the order of 10 ms), so achieving the best possible performance (one step) did not necessarily mean the listener had the ability to discriminate subtler differences in temporal envelope or good discrimination for other types of temporal cues. Indeed, many of the temporal aspects of speech can be degraded by hearing loss, including forward masking, gap detection, and temporal fine structure (see Reed, Braida, & Zurek, 2009, for a review). On the other hand, envelope has been postulated to be relatively robust to degradations caused by hearing damage. Even listeners with substantive impairments such as listeners with auditory neuropathy spectrum disorder (Wang et al., 2016) or who use auditory brainstem implants (Azadpour & McKay, 2014) have relatively good envelope perception. Recent data suggest that the relative weight placed on envelope (vs. spectral cues) may change with age but not with hearing loss per se (Goupell et al., 2017; Toscano & Lansing, 2019).

Sensitivity to spectrotemporal modulation in a non-speech signal was not associated with speech identification. Other works (Bernstein et al., 2016, 2013; Mehraei et al.,

2014; Miller et al., 2018) found a strong relationship between sensitivity to nonspeech modulation and recognition of sentences in noise. It is possible that the synthetic syllables used here did not depend on modulation sensitivity to the same extent as sentence-length tests that incorporate more envelope complexity. Another possibility is that sensitivity to nonspeech modulation is only important when background noise is present, requiring listeners to extract speech information during brief decreases in noise level and/or to segregate noise modulation patterns (Stone, Füllgrabe, & Moore, 2012).

Clinical Implications of the Cue Profile

Most clinical evaluations are centered on the pure-tone audiogram, supplemented with monosyllabic word recognition in quiet or—less often—by sentence recognition in noise (Clark, Huff, & Earl, 2017). The pure-tone audiogram is also the common basis for selection of hearing aid parameters by inputting the pure-tone audiogram into fitting software via the manufacturer's default parameters (Anderson, Arehart, & Souza, 2018). For listeners with mild-to-moderate sensorineural loss, the fact that the amount of hearing loss was not associated with the cue profile underscores the need to characterize individual abilities in a more nuanced way than can be captured by the pure-tone audiogram. Other authors have advocated for clinical batteries that characterize patient perceptual (and cognitive) abilities in more detail (Rönnerberg et al., 2016; van Esch et al., 2013). The tests described here might provide additional information as to how individual listeners use suprathreshold cues to identify phonemes.

Understanding how different listeners use suprathreshold speech information may also provide an opportunity to customize auditory treatment to listening profiles. Our preliminary data suggest that listeners who depend more heavily on temporal envelope as a consequence of poor discrimination of formant transitions demonstrate greater susceptibility to reverberation (Souza et al., 2018) and are more negatively affected by fast-acting wide dynamic range compression (Souza, Gallun, Wright, Marks, & Zacher, 2019), both speech manipulations that alter temporal envelope. It is also possible that the auditory damage that impairs formant discrimination has broad consequences and that listeners who can discriminate formants have lesser (or different) damage patterns that preserve those abilities. However, more work is needed to fully understand the circumstances under which the cue profile can offer information to guide treatment choices.

Acknowledgments

This work was supported by National Institutes of Health Grants R01 DC006014 (PI: Souza) and R01 DC015051 (PI: Gallun). The authors thank Kendra Marks for her assistance with data collection and management, Lauren Balmert for assistance with the data analysis, and Marco Rosado for technical support.

References

- Anderson, M. C., Arehart, K., & Souza, P. E. (2018). Survey of current practice in the fitting and fine-tuning of common signal-processing features in hearing aids for adults. *Journal of the American Academy of Audiology*, 29, 118–124.
- Angelo, K., & Behrens, T. (2014). *Using Speech Guard E to improve speech recognition*. Retrieved from <https://www.oticon.com/-/media/oticon-us/main/download-center/white-papers/39502-sensei-speech-guard-e-white-paper-2014.pdf>
- Azadpour, M., & McKay, C. M. (2014). Processing of speech temporal and spectral information by users of auditory brainstem implants and cochlear implants. *Ear and Hearing*, 35, e192–e203.
- Bernstein, J. G., Danielsson, H., Hällgren, M., Stenfelt, S., Rönnerberg, J., & Lunner, T. (2016). Spectrotemporal modulation sensitivity as a predictor of speech-reception performance in noise with hearing aids. *Trends in Hearing*, 20. <https://doi.org/10.1177/2331216516670387>
- Bernstein, J. G., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., & Leek, M. R. (2013). Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *Journal of the American Academy of Audiology*, 24, 293–306.
- Boothroyd, A., Springer, N., Smith, L., & Schulman, J. (1988). Amplitude compression and profound hearing loss. *Journal of Speech and Hearing Research*, 31, 362–376.
- Clark, J. G., Huff, C., & Earl, B. (2017). Clinical practice report card: Are we meeting best practice standards for adult hearing rehabilitation. *Audiology Today*, 29, 14–25.
- Davies-Venn, E., Nelson, P., & Souza, P. (2015). Comparing auditory filter bandwidths, spectral ripple detection, spectral ripple discrimination and speech recognition: Normal and impaired hearing. *The Journal of the Acoustical Society of America*, 138, 492–503.
- Dubno, J. R., Eckert, M. A., Lee, F.-S., Matthews, L. J., & Schmiedt, R. A. (2013). Classifying human audiometric phenotypes of age-related hearing loss from animal models. *Journal of the Association for Research in Otolaryngology*, 14, 687–701.
- Gallun, F. J., Seitz, A., Eddins, D. A., Molis, M. R., Stavropoulos, T., Jakien, K. M., . . . Srinivasan, N. (2018). Development and validation of Portable Automated Rapid Testing (PART) measures for auditory research. *Proceedings of Meetings on Acoustics*, 33. <https://doi.org/10.1121/2.0000878>
- Glasberg, B. R., & Moore, B. C. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America*, 79, 1020–1033.
- Goupell, M. J., Gaskins, C. R., Shader, M. J., Walter, E. P., Anderson, S., & Gordon-Salant, S. (2017). Age-related differences in the processing of temporal envelope and spectral cues in a speech segment. *Ear and Hearing*, 38, e335–e342.
- Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hoover, E. C., Souza, P. E., & Gallun, F. J. (2017). Auditory and cognitive factors associated with speech-in-noise complaints following mild traumatic brain injury. *Journal of the American Academy of Audiology*, 28, 325–339.
- Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America*, 67, 971.
- Kochkin, S. (2007). MarkeTrak VII: Obstacles to adult non-user adoption of hearing aids. *Hearing Journal*, 60, 24–50.
- Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after “temporary” noise-induced hearing loss. *Journal of Neuroscience*, 29, 14077–14085.

- Lachenbruch, P. A., & Goldstein, M. (1979). Discriminant analysis. *Biometrics*, 35, 69–85.
- Larsby, B., & Arlinger, S. (1999). Auditory temporal and spectral resolution in normal and impaired hearing. *Journal of the American Academy of Audiology*, 10, 198–210.
- Leek, M. R., Dorman, M. F., & Summerfield, Q. (1987). Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 81, 148–154.
- Leek, M. R., & Summers, V. (1996). Reduced frequency selectivity and the preservation of spectral contrast in noise. *The Journal of the Acoustical Society of America*, 100, 1796–1806.
- Lowenstein, J. H., & Nittrouer, S. (2015). All cues are not created equal: The case for facilitating the acquisition of typical weighting strategies in children with hearing loss. *Journal of Speech, Language, and Hearing Research*, 58, 466–480.
- McCormack, A., & Fortnum, H. (2013). Why do people fitted with hearing aids not wear them. *International Journal of Audiology*, 52, 360–368.
- Mehraei, G., Gallun, F. J., Leek, M. R., & Bernstein, J. G. (2014). Spectrottemporal modulation sensitivity for hearing-impaired listeners: Dependence on carrier center frequency and the relationship to speech intelligibility. *The Journal of the Acoustical Society of America*, 136, 301–316.
- Miller, C. W., Bernstein, J. G. W., Zhang, X., Wu, Y., Bentler, R. A., & Tremblay, K. (2018). The effects of static and moving spectral ripple sensitivity on unaided and aided speech perception in noise. *Journal of Speech, Language, and Hearing Research*, 61, 3113–3126.
- Moberly, A. C., Lowenstein, J. H., Tarr, E., Caldwell-Tarr, A., Welling, D. B., Shahin, A. J., & Nittrouer, S. (2014). Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing. *Journal of Speech, Language, and Hearing Research*, 57, 566–582.
- Molis, M. R., & Leek, M. R. (2011). Vowel identification by listeners with hearing impairment in response to variation in formant frequencies. *The Journal of the Acoustical Society of America*, 54, 1211–1223.
- Moore, B. C., Vickers, D. A., Plack, C. J., & Oxenham, A. J. (1999). Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism. *The Journal of the Acoustical Society of America*, 106, 2761–2778.
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatric Society*, 53, 695–699.
- Nittrouer, S., Lowenstein, J. H., Wucnich, T., & Tarr, E. (2014). Benefits of preserving stationary and time-varying formant structure in alternative representation of speech: Implications for cochlear implants. *The Journal of the Acoustical Society of America*, 136, 1845–1856.
- Pichora-Fuller, K., Schneider, B. A., Macdonald, E. N., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hearing Research*, 223, 114–121.
- Reed, C. M., Braid, L. D., & Zurek, P. M. (2009). Review of the literature on temporal resolution in listeners with cochlear hearing impairment: A critical assessment of the role of supra-threshold deficits. *Trends in Amplification*, 13, 4–43.
- Rönnerberg, J., Lunner, T., Ng, E. H., Lidestam, B., Zekveld, A., Sörqvist, P., ... Stenfelt, S. (2016). Hearing impairment, cognition and speech understanding: Exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. *International Journal of Audiology*, 55, 623–642.
- Sabin, A. T., & Souza, P. E. (2013). Initial development of a temporal-envelope-preserving nonlinear hearing aid prescription using a genetic algorithm. *Trends in Amplification*, 17, 94–107.
- Shrivastav, M. N., Humes, L. E., & Kewley-Port, D. (2006). Individual differences in auditory discrimination of spectral shape and speech-identification performance among elderly listeners. *The Journal of the Acoustical Society of America*, 119, 1131–1142.
- Souza, P., Gallun, F., Wright, R., Marks, K., & Zacher, P. (2019). Does the speech cue profile affect response to temporal distortion? Poster presented at the American Auditory Society, Scottsdale, AZ.
- Souza, P., & Hoover, E. (2018). The physiologic and psychophysical consequences of severe-to-profound hearing loss. *Seminars in Hearing*, 39, 349–363.
- Souza, P., Hoover, E., Blackburn, M., & Gallun, F. J. (2018). The characteristics of adults with severe hearing loss. *The Journal of the American Academy of Audiology*, 29, 764–779.
- Souza, P., Wright, R., Blackburn, M. C., Tatman, R., & Gallun, F. J. (2015). Individual sensitivity to spectral and temporal cues in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research*, 58, 520–534.
- Souza, P., Wright, R., Gallun, F., & Reinhart, P. (2018). Reliability and repeatability of the speech cue profile. *Journal of Speech, Language, and Hearing Research*, 61, 2126–2137.
- Stebbins, W. C., Hawkins, J. E., Jr., Johnson, L. G., & Moody, D. B. (1979). Hearing thresholds with outer and inner hair cell loss. *American Journal of Otolaryngology*, 1, 15–27.
- Stevens, K. N. (1997). Articulatory-acoustic-auditory relationships. In W. J. Hardcastle & J. Laver (Eds.), *Handbook of phonetic sciences* (pp. 462–506). Oxford, United Kingdom: Blackwell.
- Stone, M. A., Füllgrabe, C., & Moore, B. C. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America*, 132, 317–326.
- Toscano, J. C., & Lansing, C. R. (2019). Age-related changes in temporal and spectral cue weights in speech. *Language and Speech*, 62, 61–79.
- Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, 34, 434–464.
- Turgeon, C., Champoux, F., Lepore, F., Leclerc, S., & Elleberg, D. (2011). Auditory processing after sport-related concussions. *Ear and Hearing*, 32, 667–670.
- van Esch, T. E. M., Kollmeier, B., Vormann, M., Lyzenga, J., Houtgast, T., Hällgren, M., ... Dreschler, W. A. (2013). Evaluation of the preliminary auditory profile test battery in an international multi-centre study. *International Journal of Audiology*, 52, 305–321.
- Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2014). Information for coarticulation: Static signal properties or formant dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1228–1236.
- Wang, S., Dong, R., Liu, D., Wang, Y., Mao, Y., Zhang, H., Zhang, L., & Xu, L. (2016). Perceptual separation of sensorineural hearing loss and auditory neuropathy spectrum disorder. *The Laryngoscope*, 126, 1420–1425.
- Weile, J. N., Behrens, T., & Wagener, K. (2011). An improved option for people with severe to profound hearing losses. *Hearing Review*, 18, 32–45.