Research Article

Reliability and Repeatability of the Speech Cue Profile

Pamela Souza,^{a,b} Richard Wright,^c Frederick Gallun,^{d,e} and Paul Reinhart^a

Purpose: Researchers have long noted speech recognition variability that is not explained by the pure-tone audiogram. Previous work (Souza, Wright, Blackburn, Tatman, & Gallun, 2015) demonstrated that a small number of listeners with sensorineural hearing loss utilized different types of acoustic cues to identify speechlike stimuli, specifically the extent to which the participant relied upon spectral (or temporal) information for identification. Consistent with recent calls for data rigor and reproducibility, the primary aims of this study were to replicate the pattern of cue use in a larger cohort and to verify stability of the cue profiles over time.

Method: Cue-use profiles were measured for adults with sensorineural hearing loss using a syllable identification task consisting of synthetic speechlike stimuli in which spectral and temporal dimensions were manipulated along continua. For the first set, a static spectral shape varied from alveolar to palatal, and a temporal envelope rise time varied from affricate to fricative. For the second set,

formant transitions varied from labial to alveolar and a temporal envelope rise time varied from approximant to stop. A discriminant feature analysis was used to determine to what degree spectral and temporal information contributed to stimulus identification. A subset of participants completed a 2nd visit using the same stimuli and procedures.

Results: When spectral information was static, most participants were more influenced by spectral than by temporal information. When spectral information was dynamic, participants demonstrated a balanced distribution of cue-use patterns, with nearly equal numbers of individuals influenced by spectral or temporal cues. Individual cue profile was repeatable over a period of several months.

Conclusion: In combination with previously published data, these results indicate that listeners with sensorineural hearing loss are influenced by different cues to identify speechlike sounds and that those patterns are stable over time.

The ability to interpret acoustic cues as meaningful phonemes is essential to our ability to communicate. For example, spectral dimensions of the signal, such as release-burst spectra, frication spectra, and formant transitions all provide cues to a consonant's place of articulation (e.g., Delattre, Liberman, & Cooper, 1955; Dorman, Studdert-Kennedy, & Raphael, 1977; Liberman, Delattre, Gerstman, & Cooper, 1956; Smits, ten Bosch, & Collier, 1996), whereas temporal aspects of the signal

such as amplitude envelope and rise time provide cues to a consonant's manner or articulation (e.g., Bailey & Summerfield, 1980; Mann & Repp, 1981). Voice onset time, a complex temporal dimension, provides cues to a consonant's voicing (e.g., Faulkner & Rosen, 1999; Lisker & Abramson, 1964). Spectral and temporal information in the signal also carried information about speech prosody that is linked to the larger meaning of an utterance (e.g., Cutler & Clifton, 1985; Grosjean & Gee, 1987).

Hearing loss impairs speech recognition by reducing or preventing audibility of the acoustic cues. However, even audible (i.e., suprathreshold) information may not be discriminated by the listener. Poor recognition may occur for various reasons, with the most likely being loss of outer hair cells resulting in broader auditory filters (Baker & Rosen, 2002; Moore, Vickers, Plack, & Oxenham, 1999). Numerous authors have suggested that broader auditory filters can impair ability to distinguish spectral details, including shape of the burst or frication spectra (Shrivastav, Humes, & Kewley-Port, 2006), distinguishing formant frequencies (Dreisbach, Leek, & Lentz, 2005; Leek, Dorman, &

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

^bKnowles Hearing Center, Northwestern University, Evanston, IL

^cDepartment of Linguistics, University of Washington, Seattle

^dNational Center for Rehabilitative Auditory Research, Portland VA Medical Center, Oregon

^eOtolaryngology–Head and Neck Surgery, Oregon Health and Science University, Portland

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

Editor-in-Chief: Julie Liss

Editor: Megan McAuliffe

Received September 9, 2017

Revision received January 13, 2018

Accepted April 8, 2018

https://doi.org/10.1044/2018_JSLHR-H-17-0341

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

Summerfield, 1987; Leek & Summers, 1996; Molis & Leek, 2011; Souza, Wright, & Bor, 2012; Turner & Holte, 1987), and following formant frequency transitions (Coez et al., 2010; Turner, Smith, Aldridge, & Stewart, 1997).

In cases where spectral resolution is impaired, it is presumed that the listener will rely to a relatively greater extent on temporal information. Perhaps the most salient example of this is research on listeners with cochlear implants, whose devices cannot represent spectral information with high fidelity. Cochlear implant users typically cannot perceive fine-grained spectral information (Moberly et al., 2014; Winn & Litovsky, 2015). As a consequence of their impaired spectral resolution, they also experience considerable difficulty listening in noise (e.g., Wilson & Dorman, 2008).

The use of spectral versus temporal cues has not been well described for listeners with hearing loss. In general, previous work has indicated that listeners with sensorineural hearing loss are relatively less sensitive to spectral cues than to temporal cues (e.g., Hedrick & Younger, 2001, 2003, 2007; Lindholm, Dorman, Taylor, & Hannley, 1988). One limitation of such work is that it considers the effects of hearing loss for a group. Such data, while informative, are of limited use in treatment decisions for individual listeners. If we assume that all listeners with a given audiometric profile rely on similar information when it comes to speech cue access, we may overlook alternative approaches for individual listeners who respond differently. Those differences may have consequences for treatment. For example, a number of authors have advocated preserving fidelity of temporal envelope information for listeners who rely on those cues (Angelo & Behrens, 2014; Sabin & Souza, 2013; Weile, Behrens, & Wagener, 2011). Other work has suggested that some listeners with sensorineural loss will benefit from spectral enhancement (DiGiovanni, Nelson, & Schlauch, 2005). However, such manipulations require that we first identify the cue profile of an individual.

Only a small number of studies have examined individual use of spectral versus temporal cues among listeners with similar amounts of hearing loss. As one example, Hedrick and Younger (2001) reported data for three participants with hearing loss who demonstrated very different patterns of responses to synthetically modified speech cues: one was able to use both amplitude envelope and formant transition; one was insensitive to the formant transition; and a third did not appear to be sensitive to alteration of either cue. In another example (Wang & Humes, 2008), some listeners with hearing loss who responded poorly to formant enhancement (i.e., to enhancement in the spectral domain) were found to be relying to a greater extent on temporal cues. Most recently, Winn et al. (Winn, Won, & Moon, 2016) conducted a study with 19 postlingually deaf cochlear implant users and 10 listeners with normal hearing comparing speech-cue stimuli to psychoacoustic stimuli. The speech-cue stimuli consisted of a spectral manipulation of formant transitions in a /b ~ d/ continuum and temporal manipulation of voice onset time in a voiced to voiceless continuum (/p ~ b/ or /t ~ d/). Winn et al. found that spectral

ripple was most closely correlated with formant transitions and that the speech-cue stimuli more closely correlated with word recognition than any of their psychoacoustic tests. However, the exploration of individual cue-use patterns was incidental to the main experimental questions of those studies and was thus not fully examined.

A recent study (Souza et al., 2015) explored this issue in more detail using a matrix of synthetic speechlike consonant–vowel syllables in which spectral and temporal cues were manipulated along continua. As one example, temporal envelope rise times from long (as in fricatives /ʃ/ or /s/) to short (as in affricates /tʃ/ or /ts/) were combined with a static spectral energy ranging from an alveolar (e.g., /s/) to a palatal (e.g., /ʃ/) representation. The premise was that a listener who was insensitive to spectral information would identify different sounds along the envelope rise time continuum, but not along the spectral energy continuum. In general, the findings were as follows: (a) Most (but not all) listeners with hearing loss could utilize differences in static spectral shape, specifically the frequency content of the /s/-/ʃ/ contrast; (b) when the spectral cue was a formant that varied over time, only a minority of listeners with hearing loss were able to use that cue for identification; (c) among the minority of listeners who utilized formant transitions, there was a tendency for better high-frequency pure-tone thresholds.

Although compelling, the original dataset sampled a small number of participants with hearing loss ($n = 10$ for some contrasts and $n = 9$ for others). Researchers and administrators at the National Institutes of Health and elsewhere have encouraged reproducibility of preclinical data (Collins & Tabak, 2014; Downing, 2004; McIntosh et al., 2017; Mobley, Linder, Braeuer, Ellis, & Zwelling, 2013). The American Psychological Association has also affirmed the value of data replication to science (e.g., Pashler & Wagenmakers, 2012). Consistent with recent calls for data rigor and reproducibility, the first aim of this study was to replicate the pattern of cue use in a larger cohort of listeners with sensorineural hearing loss.

In addition to understanding the distribution of cue-use patterns among listeners with sensorineural hearing loss, it is also important to ensure that those profiles are stable. The profile is unlikely to be of value in making rehabilitative decisions if it changes over time. Previous studies, which assessed use of spectral and/or temporal cues to speech (e.g., Hedrick & Younger, 2001, 2007; Souza et al., 2015), have sampled only a single timepoint, leaving open the question as to whether listeners maintain a consistent profile over time. Accordingly, the second aim of this study was to demonstrate the stability of the cue profile over time.

Method

Participants

Participants included 22 adults (10 female) aged 60–89 years (mean age = 74 years) with mild to moderately severe

sensorineural hearing loss. Eight of 22 listeners were current hearing aid users. All of the participants had symmetrical loss (defined as between-ear pure-tone averages [0.5, 1, 2 kHz] $\leq$ 10 dB), and one ear was selected for testing. Audiograms for the test ears are shown in Figure 1. Mean pure-tone average was 34 dB HL (range = 17–53 dB HL). The audiograms of the participants in the present dataset were similar to those of Souza et al. (2015), where participants had a mean pure-tone average of 28 dB (range = 17–40 dB HL).

To assess reliability of the cue profile (i.e., whether similar cue-use patterns would be demonstrated on repeat testing), 10 of the participants (four female) participated in a second session (described in more detail below). Those 10 participants ranged in age from 60 to 89 years, with a mean age of 76 years. The audiometric profile was representative of the larger sample (mean pure-tone average = 36 dB HL, range = 23–48 dB HL).

Materials

Stimuli were the same as used by Souza et al. (2015) and consisted of speechlike consonant–vowel tokens synthesized using the Synthworks implementation of the Klatt parametric synthesizer. The stimuli were based on published values by Klatt (1980) and modified with the manipulation of the parameter values to create natural-sounding stimuli that were nevertheless identical on either the spectral or temporal dimensions at the endpoints. Two stimulus sets were created, each resulting in a two-dimensional spectral by a temporal matrix in which the midpoints are ambiguous either spectrally or temporally. In both stimulus sets, keeping the endpoints identical on the temporal or spectral dimension naturally distances the stimuli from naturally occurring equivalent speech sounds. For example, in naturally occurring speech, the alveolar (/s/) and palatal (/ʃ/) sounds have distinct spectro-temporal dynamics in their

Figure 1. Individual test ear audiograms (thin lines); group mean audiograms (thick lines).

main peak, in addition to overall differences in duration, that a listener might use to classify them (Reidy, 2016).

For Stimulus Set 1, the signal spectrum was varied in five steps between a distribution of energy between 3500 and 10000 Hz, with a peak at 4500 Hz (typical of /s/), and a distribution of energy between 1000 and 5500 Hz, with a peak at 3500 Hz (typical of /ʃ/). For each step in the spectral continuum, the rise time and duration of the frication noise was varied in five steps from a 90-ms rise time and a 190-ms fricative noise duration (typical of /s/ or /ʃ/) to a 2-ms rise time and a 102-ms fricative noise duration (typical of /ts/ or /tʃ/). In all tokens of Stimulus Set 1, the formant frequencies and bandwidths of the vowel portion of the syllable were typical of the vowel /i/, spoken by a male talker. The result was a matrix of 25 different tokens, illustrated schematically in Figure 2. The four “endpoint” combinations of energy distribution and frication rise time/duration represented /si/, /ʃi/, /tsi/, or /tʃi/. Along a single row of cells, the difference between stimuli was the frequency location of the spectral energy. Along a single column of cells, the difference between stimuli was the envelope rise time. Along a diagonal of cells, rise time covaried with spectral energy. In each stimulus, the listener could access either spectral or temporal aspects of the stimuli, or both, and the final percept would result from the combination of cues perceived by the listener. We hypothesized that listeners who did not utilize spectral information would demonstrate similar response patterns for tokens along a “row” as represented in Figure 2. Conversely, listeners who did not utilize temporal information would demonstrate similar response patterns for tokens along a “column” as represented in Figure 2.

Stimulus Set 2 used a similar scheme but varied formant frequencies, formant transition time, and envelope rise time (see Figure 3). The formant frequencies were varied dynamically in five steps. At the alveolar extreme, F2 onset was 1500 Hz, and F3 onset was 2700 Hz, typical of /l/. At the labial extreme, F2 onset was 850 Hz, and F3 onset was 2150 Hz, typical of /w/. The remaining formant

Figure 2. Schematic representation of dimensions manipulated in Stimulus Set 1.

Figure 3. Schematic representation of dimensions manipulated in Stimulus Set 2. The gray diagonal line refers to the stimuli examples shown in Figure 4.

values were fixed, such that F1 was 250 Hz and F4 was 3250 Hz. Amplitude rise time and formant transition time were varied in five steps. At one extreme, rise time was 60 ms with a 110-ms formant transition (typical of approximants /l/ or /w/). At the other extreme, rise time was 10 ms with a 40-ms formant transition (typical of stops /b/ or /d/). In all tokens of Stimulus Set 2, the formant frequencies and bandwidths of the vowel portion of the syllable were typical of the vowel /a/, spoken by a male talker. The four endpoint combinations of energy distribution and frication rise time/duration thus represented speechlike syllables /wa/, /la/, /ba/, or /da/. Figure 4 shows schematized examples of selected continua along the diagonal /d/ to /w/ continuum (indicated by the gray diagonal line in Figure 3).

Procedure

Testing was conducted in a double-walled, sound-treated booth. The stimuli were converted from digital to analog format (TDT RX6) and passed through a programmable attenuator (TDT PA5) and headphone buffer (TDT HB7) for presentation via an ER-2 insert earphone to the participant's test ear. Frequency shaping was deliberately not used to maintain the spectral and temporal properties of the signals as designed. Rather, to compensate for hearing loss, a flat individual gain was applied for each listener. The aim of the individual gain was to provide audible bandwidth through 5 kHz. A re-analysis of data from Souza et al. (2015) indicated that interpolation at 5 kHz produced nearly identical values to a customized (nonaudiometer) 5-kHz threshold measurement. Accordingly, pure-tone thresholds for 4 and 6 kHz were averaged to obtain the predicted 5-kHz threshold, and the stimulus presentation level was set to 10 dB SL with regard to that value or to a minimum value of 65 dB SPL, whichever was higher. The same presentation level was used for both stimulus sets.

Participants used a computer mouse to select which stimulus they heard from a graphical display of four alternatives, labeled orthogonally as either *see*, *shee*, *chee*, *tsee*

(for Stimulus Set 1) or *dah*, *lah*, *bah*, *wah* (for Stimulus Set 2). Stimulus Set 1 was tested first, followed by Stimulus Set 2. In each set, participants completed the following tasks. Participants completed 40 familiarization trials in which an endpoint was played and the correct alternative was automatically highlighted green. The participant confirmed their recognition of the label by selecting it using a computer mouse. Next, participants completed training trials, in which a randomly selected endpoint was played and the participant was asked to identify that token from the set of possible alternatives. Training was organized into blocks of 40 trials, each consisting of 10 presentations of each of the four endpoints, in random order. A visual correct answer feedback was provided after each response. After each block, the percent correct was measured for each endpoint. If the participant scored at least 80% correct for each endpoint, the training was discontinued and the listener proceeded to the testing phase, but if not, additional training blocks were presented, up to a total of four blocks (160 total trials). Finally, the participant completed 375 testing trials consisted of randomly ordered presentations (15 trials of each token, including both endpoint tokens and intermediate tokens) without correct answer feedback. During testing, rest breaks were given after 125 and 250 trials and between the two stimulus sets.

A subset of 10 participants returned for a later test session using the same paradigm as described above. The timing of the second visit varied depending on participant schedules, with a mean between-visit interval of 58 days (range = 31–82 days). All of the returning participants reported no change in perceived hearing ability or significant change in otological or medical history since the first study visit.

Results

The first aim of this study was to replicate the pattern of cue use in a larger cohort of listeners with sensorineural hearing loss. As in the original dataset (Souza et al., 2015), the likelihood of a response depended on the spectral and/or temporal dimensions of the presented syllable. This is illustrated in Figures 5 and 6, with data presented as stacked histograms. The five panels represent the five-step variation in temporal aspects of the signal. Within a panel, the x-axis represents the five-step variation in a signal spectrum. Accordingly, each individual bar represents a presented token. Participant responses to that token are shown as different shading, with the height of the shaded portion representing the total number of responses. For example, in Figure 5, for a stimuli with a gradual rise time (mimicking a fricative), responses are dominated by /s/ and /ʃ/. For a stimuli with a shorter rise time (mimicking an affricate), responses are dominated by /ts/ and /tʃ/. Within a temporal step (i.e., within a panel in Figure 5), responses varied according to the spectral content of the signal. For example, the number of /s/ responses decreased (and the number of /ʃ/ responses increased) as the spectral peak and bandwidth were shifted to the higher frequencies typical of a palatal

Figure 4. Parameters used to create a continuum from alveolar to labial on the spectral dimension and from stop to approximant on the temporal dimension resulting in a /d/ (d1s1) to /w/ (d5s5) continuum. The stimuli represent the steps along the diagonal indicated by the gray diagonal line in Figure 3.

Figure 5. The five panels represent the discrete steps along a continua of temporal information for Stimulus Set 1. Within each panel, the x-axis represents the discrete steps along continua of spectral information. Each individual bar represents a presented token. Participant responses to that token are shown as different shading, with the height of the shaded portion representing the total number of responses.

consonant. Figure 6 shows such patterns for Stimulus Set 2. Note that Figures 5 and 6 represent data for the entire group of 22 participants. Individuals showed different patterns, according to their perception.

Because both the independent (spectral and temporal parameter values) and dependent measures (stimulus identification) represented discrete steps along a continuum, a nonparametric analysis was appropriate. The data were analyzed using a linear discriminant analysis (Lachenbruch & Goldstein, 1979), a method attempting to maximize the

separation between groups using predictors of interest. This type of analysis identifies linear combinations of predictors that best discriminate between groups. In this case, the predictors were the temporal and spectral manipulations (i.e., the vertical and horizontal dimensions from Figures 2 and 3), and the groups were the four possible stimulus responses. The output of the analysis was a set of equations, or discriminant functions, which indicated to what degree each of the input variables contributes to the classification and how robust that classification was. Linear discriminant

Figure 6. The five panels represent the discrete steps along continua of temporal information for Stimulus Set 2. Within each panel, the x-axis represents the discrete steps along continua of spectral information. Each individual bar represents a presented token. Participant responses to that token are shown as different shading, with the height of the shaded portion representing the total number of responses.

Figure 7. Coefficient weights derived from the linear discriminant analysis for 2015 (open circles) and 2017 data (filled triangles) for Stimulus Set 1. Most participants were influenced by spectral shape cues, with a minority influenced by temporal envelope rise time information.

analysis (LDA) analysis was completed for each individual participant's data, for each of the stimulus sets. The output of the LDA for each participant's data was two equations, such that the first equation (LD1) provided the greatest discrimination, or separation, between the four stimulus choices and that the second (LD2) provided the second greatest discrimination among the four stimulus choices. Each discriminant function provided a coefficient or "weight" for each variable, reflecting its contribution or relative importance (Hair, Black, Babin, & Anderson, 2010). The larger the weight, the more important that variable's role is in the discriminant function. The LDA also provides the proportion of between-group variance accounted for by the given equation. The weights were then extracted. Thus, the *weight* here refers to the results of the model that describes the entire response pattern for an individual listener and not to a direct acoustic manipulation. The standardized coefficients mean that the variables are all measured on the same scale, allowing interpretation of relative importance of the spectral and temporal dimensions.

Figure 7 shows the output of the LDA for Stimulus Set 1, plotted as temporal weight versus spectral weight. Filled triangles show individual weights for the present dataset. For comparison, results from Souza et al. (2015) are plotted as open circles. To aid data interpretation, the diagonal line demarcates the point where, according to the LDA output, the temporal and spectral aspects of the stimuli have similar influence on a participant's response patterns. Results from this study closely replicate those of the 2015

dataset. In both cases, nearly all of the participants were influenced by manipulation of the spectral cue (i.e., spectral energy of the frication noise) for identification. Only a small minority (approximately 10% of participants in each dataset) was influenced to a greater extent by manipulation of the temporal information (i.e., rise time).

Figure 8 shows the results of the LDA for Stimulus Set 2. As in Figure 7, filled triangles show individual weights for the present dataset. For comparison, results from Souza et al. (2015) are plotted as open circles. In the present data, approximately equal numbers of participants were more influenced by each type of cue. The proportion of participants utilizing spectral information was considerably higher than in Souza et al. (2015), despite the current participants being similar to the 2015 cohort in terms of audiogram, age, and hearing loss etiology.

Repeatability of the Cue Profile

The second aim of this study was to demonstrate stability of the cue profile over time. Figures 9 and 10 show results of the LDA for the two visits. Data for each individual are plotted using the same shape-color combination. We hypothesized that listeners who had relatively greater weights for spectral dimensions would maintain that pattern across visits and that listeners who had relatively greater weights for temporal dimensions would maintain that pattern across visits. That was the case for all listeners for Stimulus Set 1 (see Figure 9) and nine of 10 listeners for Stimulus

Figure 8. Coefficient weights derived from the linear discriminant analysis for 2015 (open circles) and 2017 data (filled triangles) for Stimulus Set 2. Participants showed a more even distribution of cue use than for Stimulus Set 1, particularly in the 2017 dataset.

Figure 9. Coefficient weights derived from the linear discriminant analysis illustrating repeated testing for 10 participants for Stimulus Set 1. Results for the two test sessions for each participant are shown by a unique shape–color combination.

Figure 10. Coefficient weights derived from the linear discriminant analysis illustrating repeated testing for 10 participants for Stimulus Set 2. Results for the two test sessions for each participant are shown by a unique shape–color combination.

Set 2 (see Figure 10). One listener appeared to shift from greater use of temporal information to greater use of spectral information in the Stimulus Set 2 data (see Figure 10, filled circles). However, that listener had data that were not well fit by the LDA (proportion of variance explained was $< 50\%$ for each dimension). The remaining listeners in Stimulus Set 2 and all listeners' data for Stimulus Set 1 had substantially better fits ($\geq 70\%$ of variance).

To compare data points across visits, the data points were converted to vectors (v) as follows. The vector magnitude was calculated as follows:

$$v = \sqrt{(\text{temporal weight}^2 + \text{spectral weight}^2)}. \quad (1)$$

The vector direction (θ) was calculated as follows:

$$\theta = \tan^{-1}(\text{temporal weight}/\text{spectral weight}). \quad (2)$$

Under this scheme, the vector direction quantified the cue profile, with spectrally dependent profiles indicated by $\theta < 45^\circ$ and temporally dependent profiles indicated by $\theta > 45^\circ$. The vector magnitude quantified the strength of the weighting factor. Because the small sample was not normally distributed, a nonparametric analysis was used.

For Stimulus Set 1 (/si/-/ʃi/-/tʃi/-/tʃi/), a Wilcoxon signed-ranks analysis indicated that there was no difference in vector magnitude ($T = 23$, $p = .646$, $r = .103$) or direction

($T = 34$, $p = .508$, $r = .148$). That is, there was no significant difference in the location of the data points between the first and second visit. Similarly, for Stimulus Set 2 (/da/-/ba/-/la/-/wa/), a Wilcoxon signed-ranks analysis indicated that there was no difference in vector magnitude ($T = 46$, $p = .059$, $r = .422$) or direction ($T = 32$, $p = .646$, $r = .103$) between the first and second visits.

Influence of Degree of Hearing Loss

In the 2015 dataset, there was a trend such that listeners with relatively worse hearing thresholds tended to rely upon temporal cues rather than dynamic spectral cues. Such a finding might occur if there was a simplistic relationship whereby greater amounts of hearing loss resulted in greater loss of hair cells (Ryan & Dallos, 1975; Stebbins, Hawkins, Johnson, & Moody, 1979) and, therefore, broader auditory filters (Dallos & Harris, 1978) and that listeners with broader auditory filters were unable to track the fine-scale changes necessary to obtain dynamic formant information. However, there is likely to be a more complex relationship between pure-tone thresholds, auditory filter width, and suprathreshold recognition. Following this theme, we undertook an analysis of the relationship between the degree of loss and the cue profile. Because of the small number of temporal-weighting participants in Stimulus Set 1, only Stimulus Set 2 was analyzed. The analysis included the present cohort of 22 participants plus the 10 participants in the 2015 data. Data points were converted to vectors (v) as described above, such that vector direction

quantified the relative use of spectral versus temporal cues, whereas vector magnitude quantified the extent of reliance on the cue indicated by vector direction. Bivariate correlational analyses indicated that there was no relationship between the overall amount of hearing loss as quantified by the three-frequency pure-tone average (0.5, 1, 2 kHz) and either vector direction ($r = .16$, $p = .40$) or magnitude ($r = -.04$, $p = .85$). There was a modest correlation between the amount of high-frequency loss as quantified by high-frequency pure-tone average (2, 3, 4 kHz) and vector direction ($r = .46$, $p = .01$), but not vector magnitude ($r = .08$, $p = .66$).

Discussion

In Souza et al. (2015) and in this study, listeners demonstrated different cue profiles despite having similar pure-tone thresholds. When spectral information was relatively static and broad in frequency (i.e., spectral shape), nearly all listeners (29 of 32 across the two datasets) could access the cue. However, when spectral information was dynamic and fine grained (i.e., formant frequency), listeners were nearly evenly split in their relative ability to reliably use the formant cue or the temporally based (envelope rise time) cue.

The present data agree with previous work suggesting that sensorineural hearing loss may affect the use of formant transitions, perhaps through smearing of the formant information represented by the auditory system (Coez et al., 2010; Turner et al., 1997). However, our data also suggest that this is not an inevitable consequence of pure-tone threshold elevation. There was only a modest relationship with high-frequency threshold elevation, explaining approximately 20% of the variance in the cue profile. Perhaps the cue profile is related to changes to auditory processing that are only loosely correlated with threshold. For example, broadened auditory filters are likely to impact the use of fine-scale spectral cues, but an auditory filter width does not show a strong relationship to a pure-tone threshold (e.g., Glasberg & Moore, 1986; Souza et al., 2012). Additionally, it is unclear whether those listeners who demonstrated a temporal-reliant cue profile did so because auditory pathology limited the access to alternative cues (as with broadened auditory filters degrading fine-scale spectral cues, such as the formant transitions in Stimulus Set 2) or as a bias toward a particular cue. There is evidence for both views. On one hand, it is quite reasonable to expect that listeners with broad auditory filters would have a reduced ability to track fine-scale spectral details (Preminger & Wiley, 1985), such as the formant frequencies employed here. On the other hand, some data suggest that auditory filter width is a poor predictor of suprathreshold cue use (Bernstein, Summers, Grassi, & Grant, 2013; Summers & Leek, 1992). Other studies have shown that cue use can be shifted if a particular (and more reliable) cue becomes more uncertain (e.g., Repp, 1982; Toscano & McMurray, 2010; Webster, 2002). If the cue profile represents a bias rather than an inability, it may be possible to enhance or train reliance on a more useful cue.

The stability of the cue profile over time is important if this measure is to be used as a basis for understanding speech recognition abilities of listeners with hearing loss and, potentially, as a basis for treatment decisions. These data demonstrate the stability of the cue profile over time. The single listener who appeared to shift profiles was the sole listener with a low percentage of variance explained by the LDA weights, suggesting that the LDA model did not provide a good fit to her data. The remaining listeners had performance patterns that were statistically similar to their first test. These data are consistent with the work by Calundruccio and Doherty (2008), who measured individual weights for perturbed speech spectrum in two listeners with hearing loss, one with a 3-month test interval and the other with a 6-month test interval. In each case, there was no significant difference in the listener's weighting of specific spectral bands.

Implications

For both clinical audiologists and translational researchers, it is of interest to understand the implications of different cue-use profiles. First, these profiles may have consequences for unaided speech recognition. For example, cochlear implant users who have relatively greater sensitivity to dynamic spectral cues demonstrate better word recognition than listeners who do not utilize such cues (Moberly, Lowenstein, & Nittrouer, 2015). Such information would be useful to anticipate communication difficulties and in counseling. When the essential information is more readily supplied by spectral cues, reliance on temporal cues might be inferior to reliance on spectral cues (e.g., Xu, Thompson, & Pfingst, 2005). In such circumstances, we might expect listeners with a more temporal-reliant profile to perform more poorly on a general speech task and, especially, one in which their preferred cue is restricted. In support of such data, Figure 11 shows data from a subset of listeners in this study ($n = 12$) who also participated in a separate study of speech in reverberant conditions (Reinhart & Souza, 2018), an acoustic distortion with greater effect on temporal cues than on spectral cues. The listeners who demonstrated greater reliance on temporal cues show poorer ability to understand speech in mild levels of reverberation. While that dataset is too small to draw definitive conclusions and was not designed to directly measure this issue, it suggests that temporally distorting situations may have disproportionately negative impacts on listeners who rely more heavily on temporal cues in the speech signal. It is important to note, however, that a "spectral is better" argument cannot be universally applied. In some cases, the meaningful cue is temporal. In addition to the manner distinctions are examined in this study, categorization of voice onset time, final consonant voicing, and vowel or consonant duration all rely on temporal differences and might be well perceived by listeners with a more temporal-reliant profile.

A potential goal following the work presented here is to apply cue profiles to treatment decisions. Current digital hearing aids employ a number of features that alter

Figure 11. Preliminary data comparing the recognition of sentences in reverberation, from Reinhart and Souza (2018). Listeners are grouped by their cue profile (i.e., either spectral or temporal cue use). The center line shows the median; the box length shows the interquartile range. Whiskers show minimum and maximum, excluding outliers. The circle shows an outlier exceeding 1.5 times the interquartile range, and the asterisk shows an outlier exceeding three times the interquartile range.

both the temporal and spectral properties of speech. It might be possible to select or adjust signal-processing strategies, such that the most useful cues for an individual are preserved, or even enhanced. To take a simplistic possibility, listeners who are influenced by temporal cues might be detrimentally affected by fast-acting, wide-dynamic range compression amplification that modifies the signal envelope, and might benefit from temporal envelope-preserving amplification. Although speculative, such possibilities are well within the technical capabilities of hearing aid processors and present an interesting avenue for future study.

Acknowledgments

This work was supported by National Institutes of Health Grant NIH R01 DC006014 (awarded to P. Souza). The research reported in this publication was supported, in part, by the National Center for Advancing Translational Sciences Grant UL1TR001422, which supports the Northwestern Biostatistics Cores. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank Laura Mathews and Rachel Ellinger for their assistance in data collection and Lauren Balmert for her guidance regarding data analysis.

References

- Angelo, K., & Behrens, T. (2014). Sensei Pro-Using Speech Guard E to improve speech recognition. Oticon White Paper-Paediatric Knowledge Brief. Retrieved from <https://www.oticon.global/-/media/oticon%20us/main/download%20center/white%20papers/39502%20sensei%20speech%20guard%20e%20white%20paper%202014.pdf>
- Bailey, P. J., & Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology*, 6, 536–563.
- Baker, R. J., & Rosen, S. (2002). Auditory filter nonlinearity in mild/moderate hearing impairment. *The Journal of the Acoustical Society of America*, 111, 1330–1339.
- Bernstein, J. G., Summers, V., Grassi, E., & Grant, K. W. (2013). Auditory models of suprathreshold distortion and speech intelligibility in persons with impaired hearing. *Journal of the American Academy of Audiology*, 24, 307–328.
- Calundann, L., & Doherty, K. A. (2008). Spectral weighting strategies for hearing-impaired listeners measures using a correlational method. *The Journal of the Acoustical Society of America*, 123, 2367–2378.
- Coez, A., Belin, P., Bizaguet, E., Ferrary, E., Zibovicius, M., & Samson, Y. (2010). Hearing loss severity: Impaired processing of formant transition duration. *Neuropsychologica*, 48, 3057–3061.
- Collins, F. S., & Tabak, L. A. (2014). NIH plans to enhance reproducibility. *Nature*, 505, 612–613.
- Cutler, A., & Clifton, C. E. (1985). The use of prosodic information in word recognition. In H. Bouma & D. G. Bowhuis (Eds.), *Attention and performance X* (pp. 183–196). Hillsdale, NJ: Erlbaum.
- Dallos, P., & Harris, D. (1978). Properties of auditory nerve responses in absence of outer hair cells. *Journal of Neurophysiology*, 41, 365–383.
- Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, 27, 769–773.
- DiGiovanni, J. J., Nelson, P. B., & Schlauch, R. S. (2005). A psychophysical evaluation of spectral enhancement. *Journal of Speech, Language, and Hearing Research*, 48, 1121–1135.
- Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 22, 109–122.
- Downing, S. M. (2004). Reliability: On the reproducibility of assessment data. *Medical Education*, 38, 1006–1012.
- Dreisbach, L. E., Leek, M. R., & Lentz, J. J. (2005). Perception of spectral contrast by hearing-impaired listeners. *Journal of Speech, Language, and Hearing Research*, 48, 910–921.
- Faulkner, A., & Rosen, S. (1999). Contributions of temporal encodings of voicing, voicelessness, fundamental frequency and amplitude variation to auditory and audio-visual speech perception. *The Journal of the Acoustical Society of America*, 106, 2063–2073.
- 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.
- Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition*, 25, 135–155.
- 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.
- Hedrick, M., & Younger, M. S. (2001). Perceptual weighting of relative amplitude and formant transition cues in aided CV syllables. *Journal of Speech, Language, and Hearing Research*, 44, 964–974.
- Hedrick, M., & Younger, M. S. (2003). Labeling of /s/ and /sh/ by listeners with normal and impaired hearing, revisited. *Journal of Speech, Language, and Hearing Research*, 46, 636–648.

- Hedrick, M. S., & Younger, M. S. (2007). Perceptual weighting of stop consonant cues by normal and impaired listeners in reverberation versus noise. *Journal of Speech, Language, and Hearing Research, 50*, 254–269.
- Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America, 67*, 13–33.
- Lachenbruch, P. A., & Goldstein, M. (1979). Discriminant analysis. *Biometrics, 35*, 69–85.
- 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.
- Lieberman, A. M., Delattre, P. C., Gerstman, L. J., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology, 52*, 127–137.
- Lindholm, J. M., Dorman, M. F., Taylor, B. E., & Hannley, M. T. (1988). Stimulus factors influencing the identification of voiced stop consonants by normal-hearing and hearing-impaired adults. *The Journal of the Acoustical Society of America, 83*, 1608–1614.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–422.
- Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America, 69*, 548–558.
- McIntosh, L., Juehne, A., Hudson-Vitale, C., Liu, X., Alcoser, R., Lukas, J. C., & Evanoff, B. (2017). Repeat: A framework to assess empirical reproducibility in biomedical research. *Open Science Framework*. <https://doi.org/10.17605/osf.io/4np66>
- Moberly, A. C., Lowenstein, J. H., & Nittrouer, S. (2015). Word recognition variability with cochlear implants: “Perceptual attention” versus “auditory sensitivity.” *Ear and Hearing, 37*, 14–26.
- 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.
- Mobley, A., Linder, S. K., Brauer, R., Ellis, L. M., & Zwelling, L. (2013). A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic. *PLoS One, 8*, e63221.
- Molis, M. R., & Leek, M. R. (2011). Vowel identification by listeners with hearing impairment in response to variation in formant frequencies. *Journal of Speech, Language, and Hearing Research, 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.
- Pashler, H., & Wagenmakers, E.-J. (2012). Editors’ introduction to the special section on replicability in psychological science. *Perspectives on Psychological Science, 7*, 528–530.
- Preminger, J., & Wiley, T. L. (1985). Frequency selectivity and consonant intelligibility in sensorineural hearing loss. *Journal of Speech and Hearing Research, 28*, 197–206.
- Reidy, P. F. (2016). Spectral dynamics of sibilant fricatives are contrastive and language specific. *The Journal of the Acoustical Society of America, 140*(4), 2518–2529.
- Reinhart, P., & Souza, P. (2018). Listener factors associated with individual susceptibility to reverberation. *Journal of the American Academy of Audiology, 29*, 73–82.
- Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*, 81–110.
- Ryan, A., & Dallos, P. (1975). Effect of absence of cochlear outer hair cells on behavioural auditory threshold. *Nature, 253*, 44–46.
- 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., Humes, L., & 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.
- Smits, R., ten Bosch, L., & Collier, R. (1996). Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants, I: Perception experiment. *The Journal of the Acoustical Society of America, 100*, 3852–3864.
- Souza, P., Wright, R., Blackburn, M., 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., & Bor, S. (2012). Consequences of broad auditory filters for identification of multichannel-compressed vowels. *Journal of Speech, Language, and Hearing Research, 55*, 474–486.
- Stebbins, W. C., Hawkins, J. F. J., Johnson, L. G., & Moody, D. B. (1979). Hearing thresholds with outer and inner hair cell loss. *American Journal of Otolaryngology, 1*, 15–27.
- Summers, W. V., & Leek, M. R. (1992). The role of spectral and temporal cues in vowel identification by listeners with impaired hearing. *Journal of Speech and Hearing Disorders, 35*, 1189–1199.
- 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.
- Turner, C. W., & Holte, L. A. (1987). Discrimination of spectral-peak amplitude by normal and hearing-impaired subjects. *The Journal of the Acoustical Society of America, 81*, 445–451.
- Turner, C. W., Smith, S. J., Aldridge, P. L., & Stewart, S. L. (1997). Formant transition duration and speech recognition in normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 101*, 2822–2825.
- Wang, X., & Humes, L. (2008). Classification and cue weighting of multidimensional stimuli with speech-like cues for young normal hearing and elderly hearing-impaired listeners. *Ear and Hearing, 29*, 725–745.
- Webster, G. (2002). *Toward a psychologically and computationally adequate model of speech perception*. Seattle, WA: University of Washington.
- Weile, J. N., Behrens, T., & Wagener, K. (2011). An improved option for people with severe to profound hearing losses. *Hearing Review, 18*, 32–45.
- Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: A remarkable past and a brilliant future. *Hearing Research, 242*, 3–21.
- Winn, M., & Litovsky, R. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *The Journal of the Acoustical Society of America, 137*, 1430–1442.
- Winn, M. B., Won, J. H., & Moon, I. J. (2016). Assessment of spectral and temporal resolution in cochlear implant users using psychoacoustic discrimination and speech cue categorization. *Ear and Hearing, 37*, e377–e390.
- Xu, L., Thompson, C. S., & Pflingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *The Journal of the Acoustical Society of America, 117*, 3255–3267.