



**Individual sensitivity to spectral and temporal cues in
listeners with hearing impairment**

Journal:	<i>Journal of Speech, Language, and Hearing Research</i>
Manuscript ID:	JSLHR-H-14-0138.R2
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Souza, Pam; Northwestern University, Roxelyn and Richard Pepper Department of Communication Sciences and Disorders Wright, Richard; University of Washington, Linguistics Blackburn, Michael; Northwestern University, Communication Sciences and Disorders Tatman, Rachael; University of Washington, Linguistics Gallun, Frederick (Erick); Portland VA Medical Center, National Center for Rehabilitative Auditory Research
Keywords:	Speech recognition, Hearing loss, Spectral, Temporal

Individual sensitivity to spectral and temporal cues in listeners with hearing impairment

Pamela E. Souza^{1,2}Richard A. Wright³Michael C. Blackburn¹Rachael Tatman³Frederick J. Gallun^{4,5}¹Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL²Knowles Hearing Center, Northwestern University, Evanston, IL³Department of Linguistics, University of Washington, Seattle, WA⁴National Center for Rehabilitative Auditory Research, Portland VA Medical Center, Portland,

OR

⁵Oregon Health & Science University, Portland, OR

Individual sensitivity to spectral and temporal cues in listeners with hearing impairment

Purpose: The present study was designed to evaluate use of spectral and temporal cues, under conditions where both types of cues were available.

Method: Participants included adults with normal hearing and hearing loss. We focused on three categories of speech cues: static spectral (spectral shape); dynamic spectral (formant change); and temporal (amplitude envelope). Spectral and/or temporal dimensions of synthetic speech were systematically manipulated along a continuum and recognition was measured using the manipulated stimuli. Level was controlled to ensure cue audibility. Discriminant function analysis was used to determine to what degree spectral and temporal information contributed to the identification of each stimulus.

Results: Listeners with normal hearing were influenced to a greater extent by spectral cues for all stimuli. Listeners with hearing impairment generally utilized spectral cues when the information was static (spectral shape) and but used temporal cues when the information was dynamic (formant transition). The relative use of spectral and temporal dimensions varied among individuals, especially among listeners with hearing loss.

Conclusions: Information about spectral and temporal cue use may aid in identifying listeners who rely to a greater extent on particular acoustic cues, and to apply that information toward therapeutic interventions.

Keywords: hearing loss, speech recognition, spectral, temporal

1
2
3 Individual sensitivity to spectral and temporal cues in listeners with hearing impairment

4
5 Listeners with hearing loss often report that speech sounds distorted. Some of the
6
7
8 “distortion” is due to threshold elevation which reduces consonant audibility (Humes, 2007;
9
10 Singh & Allen, 2012). However, there are also situations where speech sounds are audible but
11
12 speech recognition is still poor (e.g., Bernstein, Mehraei, et al., 2013; Bernstein, Summers,
13
14 Grassi, & Grant, 2013; Souza, Boike, Witherell, & Tremblay, 2007). Those problems are often
15
16 attributed to a generalized problem resolving the spectral and/or temporal cues in speech.
17
18

19
20 Spectral cues to consonant place of articulation include the frequency characteristics of
21
22 the consonant release burst or frication noise in obstruents (e.g. LaRiviere, Winitz, & Herriman,
23
24 1975) and the onset frequency location of the formants and resulting formant transitions in
25
26 sonorants and obstruents (e.g. Dorman, Studdert-Kennedy, & Raphael, 1977), both of which
27
28 provide cues to consonant place. Temporal cues derived from the amplitude envelope include
29
30 the duration of the sound and the rise time at onset of the consonant. Rise time and duration
31
32 provide information about consonant manner (e.g., /ʃ/ - /tʃ/-/t/), vowel identity (e.g., /ɪ/-/i/)
33
34 (Howell & Rosen, 1983; Rosen, 1992), and consonant voicing (Stevens, Blumstein, Glicksman,
35
36 Burton, & Kurowski, 1992). It is well known that there is a one-to-many and many-to-one
37
38 relationship between speech cues and speech sounds resulting in significant informational
39
40 redundancy. That, in turn, underlies the remarkable perceptual robustness of the speech signal
41
42 (see discussions in Goldinger & Azuma, 2003; Nearey, 1997; Wright, 2004). For listeners with
43
44 normal hearing presented with everyday speech, when one aspect of the signal is distorted, the
45
46 listener may be able to shift to an alternative cue (Repp, 1983).
47
48
49
50
51

52
53 The same level of redundancy is unlikely for listeners with hearing loss. Some listeners
54
55 with hearing loss have broadened auditory filters (Faulkner, Rosen, & Moore, 1990; Glasberg &
56
57
58
59
60

1
2
3 Moore, 1986; Souza, Wright, & Bor, 2012) which could limit their ability to resolve and use
4
5 spectral cues. That limitation is supported by studies which show that listeners with hearing loss
6
7 have difficulty identifying consonants when the frequency content of the consonant falls into a
8
9 region of broadened auditory filters (Dubno, Dirks, & Ellison, 1989; Preminger & Wiley, 1985),
10
11 and have more difficulty identifying vowels based on formant patterns as compared to the
12
13 average performance of listeners with normal hearing (Leek & Summers, 1996; Molis & Leek,
14
15 2011; Souza, Wright, et al., 2012; Turner & Henn, 1989). Other studies show that formant
16
17 *transitions*—where formant frequency varies dynamically across the coarticulation point of two
18
19 sounds—may be particularly difficult for listeners with hearing loss (Carpenter & Shahin, 2013;
20
21 Coez et al., 2010; Hedrick, 1997; Hedrick & Younger, 2007; Stelmachowicz, Kopun, Mace,
22
23 Lewis, & Nittrouer, 1995; Turner, Smith, Aldridge, & Stewart, 1997; Zeng & Turner, 1990). In
24
25 most of these studies, however, there were also individuals with hearing loss who showed similar
26
27 abilities to those with normal hearing, suggesting that ability to resolve spectral cues cannot be
28
29 easily predicted from the individual audiogram.
30
31
32
33
34
35

36
37 For listeners for whom spectral cues are less accessible, it seems reasonable to expect
38
39 greater reliance on temporal cues. The idea that listeners with hearing loss will be forced to shift
40
41 reliance to temporal cues has been put forth in several papers (Boothroyd, Springer, Smith, &
42
43 Schulman, 1988; Christensen & Humes, 1997; Davies-Venn & Souza, 2014; Davies-Venn,
44
45 Souza, Brennan, & Stecker, 2009; Souza, Jenstad, & Folino, 2005). That hypothesis relies, in
46
47 part, on the idea that temporal cues will be more resistant to degradation from hearing loss than
48
49 spectral cues, provided the listener can access a sufficiently wide signal bandwidth. However,
50
51 while many individuals with hearing loss have good temporal cue perception (Reed, Braid, &
52
53 Zurek, 2009), others—particularly older listeners—demonstrate reduced temporal-cue perception
54
55
56
57
58
59
60

1
2
3 (e.g., Brennan, Gallun, Souza, & Stecker, 2013). This makes it difficult to predict the relative
4
5 use of spectral versus temporal cues with any precision.
6
7

8 Most studies of cue use by listeners with hearing loss used a single-cue paradigm; that is,
9
10 they varied one dimension (either spectral or temporal) and assessed performance on that basis
11
12 (e.g., Coez et al., 2010; Hedrick & Rice, 2000; Souza & Kitch, 2001; Turner et al., 1997). With
13
14 that approach, it is difficult to say with certainty whether a particular listener with hearing loss
15
16 listening to speech which contains *both* spectral and temporal information—as would be the case
17
18 in most everyday listening situations—will be utilizing temporal cues, spectral cues or both.
19
20 Individual characterization could provide insight into auditory damage patterns, and may also be
21
22 useful in a practical way. For example, amplification parameters could be set to avoid distorting
23
24 information useful to that individual. To illustrate this idea, consider two hypothetical listeners:
25
26 one who has relatively good spectral resolution and can utilize spectral cues to speech; and one
27
28 who has reduced spectral resolution and consequently depends to a greater extent on temporal
29
30 cues. Without this distinction, a common rehabilitation approach might be to use fast-acting
31
32 compression amplification to improve audibility of low-intensity phonemes. Because that type of
33
34 amplification would also distort temporal envelope cues (Jenstad & Souza, 2005), it is possible
35
36 that such a strategy may be detrimental to the listener who depends to a greater extent on
37
38 temporal cues (Davies-Venn & Souza, 2014). Conversely, use of a high number of compression
39
40 channels—which may improve audibility at the expense of smoothed frequency contrasts (Bor,
41
42 Souza, & Wright, 2008)—may be detrimental to the listener who depends to a greater extent on
43
44 spectral cues. Although such arguments are speculative, a necessary first step in disentangling
45
46 these issues is to understand how reliance on spectral and temporal properties varies among
47
48 individuals with hearing loss.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

It seems likely that as a group, adults with hearing loss will tend to rely more on temporal cues when both spectral and temporal cues are available (e.g., Hedrick & Younger, 2003; 2007; Lindholm, Dorman, Taylor, & Hannley, 1988). Some studies hint that individuals with hearing loss may be accessing different stimulus dimensions. Within a small sample (three participants with hearing loss), Hedrick and Younger (2001) demonstrated three different patterns of responses to synthetically-modified speech cues: one listener was sensitive to both amplitude envelope and formant transition (similar to listeners with normal hearing); a second listener was heavily influenced by amplitude envelope and insensitive to the formant transition; and a third listener did not appear to respond to either cue modification. Work with non-speech signals (Wang & Humes, 2008) found that listeners with hearing loss were able to use an enhanced formant transition to improve sound labeling. However, that study also excluded several listeners who were unable to perform the task to a criterion level. Post-hoc examination of the excluded listeners suggested that one of them used temporal cues to a greater extent than the listeners who were retained in the test group. Two of the listeners who were included in the test group but performed at low levels also showed evidence of greater temporal-cue use. Taken together, these results suggest that individuals with hearing loss may show different sensitivity to temporal versus spectral cues.

43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The present study was designed to evaluate individual use of spectral and temporal cues, under conditions where both types of cues were available. We also aimed to construct a complete set of stimuli with more dimensions (static spectral, dynamic spectral, temporal; low- and high-frequency cues) applied to the same set of individuals than have been tested in previous research. Synthetic speech stimuli were used to provide precise control over dimensions of the stimuli. We focused on three categories of speech cues: static spectral (spectral shape); dynamic

1
2
3 spectral (formant change); and temporal (amplitude envelope). Two sets of stimuli were created.
4
5 In one set, the spectral cue was the location of high-frequency frication noise, and the temporal
6
7 cue was the envelope rise-time and duration of frication. In the second set, the spectral cue was
8
9 the initial frequency of the formants and rate of formant change; the temporal cue was envelope
10
11 rise time. In Experiment 1, we confirmed sensitivity to these speech dimensions in listeners with
12
13 normal hearing. In Experiment 2, we examined variability among adults with sensorineural
14
15 hearing loss, while taking care to ensure sufficient audibility of available cues.
16
17
18
19

Experiment 1

20
21 As a foundation to our eventual work with listeners with hearing impairment, the purpose
22
23 of Experiment 1 was to demonstrate that listeners with normal hearing are sensitive to both
24
25 spectral and temporal aspects of a synthetic speech stimulus. We further hypothesized that all
26
27 listeners with normal hearing would show similar response patterns.
28
29
30
31

Method

Participants

32
33
34
35
36 Participants included 20 adults with normal hearing (15 female), aged 19-30 years (mean
37
38 age 22 years). All participants had bilaterally normal hearing, defined as thresholds of 20 dB HL
39
40 or better (ANSI, 2004) at octave frequencies between 0.25 and 8 kHz; spoke English as their first
41
42 language; completed an informed consent process and were compensated for their time.
43
44
45

Stimuli

46
47
48 Two sets of stimuli were created to test the interaction between spectral cues and
49
50 temporal cues. All stimuli were synthesized using the Synthworks implementation of the Klatt
51
52 parametric synthesizer (v. 1112) running on a Macbook Pro with Mac OSX 10.7.4 and saved as
53
54 individual sound files on a computer hard drive.
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

Stimulus set 1: We first created a set of stimuli that varied in the distribution of energy in the frication spectrum. The frequency values were based on published values in Klatt (1980)¹. At one extreme, the signal spectrum had a distribution of energy between 3500 Hz and 10000 Hz with a peak at 4500 Hz (typical of the alveolar consonant /s/; see Table 1, column 2). At the other extreme, the signal spectrum had a distribution of energy between 1000 Hz and 5500 Hz with a peak at 3500 Hz (typical of the palatal consonant /j/; see Table 1, column 6). The resulting waveforms and spectrograms are shown in the top left and lower left panels of Figure 1. Intermediate values (Table 1, columns 3-5) were chosen to span the range between the alveolar and palatal “endpoints”.

Next, we converted each unique frequency spectra (i.e., the spectra denoted by each column of Table 1) into five different tokens by manipulating the rise time² and duration of the frication noise. One end of the continuum had a 90 ms rise-time and a 190 ms fricative noise duration (typical of the fricatives /s/ or /ʃ/, as shown in Figure 1, left panels). The other end of the continuum had a 2 ms rise time and a 102 ms fricative noise duration (typical of the affricates /ts/ or /tʃ/, as shown in Figure 1, right panels). Intermediate values were chosen to span the range between the fricative and affricate “endpoints”.

In all of the stimuli the vowel portion of the syllable was typical of the vowel /i/ with the following values (where F=formant, b=formant bandwidth): duration 325 ms; F1 310 Hz; F2 2100 Hz; F3 2700 Hz; b1 45 Hz; b2 200 Hz; b3 400 Hz. The fundamental frequency (f_0) was set to 150 Hz for the initial 60 ms then fell linearly to 100 Hz. Amplitude of voicing (av) rose from

¹ To remove low frequency noise that is not characteristic of natural unvoiced fricatives, the fricative portion was high-pass filtered using a Hanning filter with a cutoff of 500 Hz and a smoothing setting of 100 Hz using Praat (Boersma, 2001).

² A linear attenuation slope was applied to the base (fricative) shortening the rise time in 22 ms steps (using Sound Studio 3.6).

1
2
3 0 to 60 over the initial 40 ms, then was 60 dB until falling to 0 dB over the final 60 ms of the
4
5 stimulus.
6

7
8 The combination of five values on spectral and five on the temporal dimension created 25
9
10 total stimuli. The combination of the most extreme values on each continuum created four
11
12 “endpoints” with spectral and temporal characteristics typical of /si, ʃi, tsi, tʃi/ (Figure 1).
13
14

15 *Stimulus set 2:* The second stimulus set was similarly designed to create a two-
16
17 dimensional continuum with spectral variation on one dimension and temporal variation on the
18
19 other. For this set, we wished to focus on use of lower-frequency cues (where participants with
20
21 hearing loss were likely to have narrower auditory filters and better spectral resolution, as well as
22
23 greater audibility). Stimulus set 2 also employed a dynamic spectral cue (formant transitions),
24
25 rather than the relatively static spectra which comprised the spectral cue for stimulus set 1.
26
27

28
29 We first created a set of stimuli that varied in F2 and F3 frequency and bandwidth onset
30
31 values. The formant frequencies (F) and bandwidths (b) were based on values published in Klatt
32
33 (1980)³. At one extreme, the formants had onset values of: F2 1500 Hz; F3 2700 Hz; b2 100
34
35 Hz; b3 725 Hz (typical of the alveolar consonant /l/; see Table 2 column 2). At the other extreme,
36
37 the formants had onset values of: F2 850 Hz; F3 2150 Hz; b2 95 Hz; b3 95 Hz (typical of the
38
39 labial consonant /w/; see Table 2, column 6). The remaining formant and bandwidth onset values
40
41 were fixed: F1 250 Hz; F4 32150 Hz; b1 55 Hz; b4 200 Hz. The resulting waveforms and
42
43 spectrograms are shown in the top and lower left panels of Figure 2. Intermediate values (Table
44
45 2, columns 3-5) were chosen to span the range between the alveolar and labial “endpoints”.
46
47
48
49

50
51 Next, we converted each unique frequency spectra (i.e., the spectra denoted by each
52
53 column of Table 2) into five different tokens by manipulating the amplitude rise time and
54
55 formant transition time. One end of the continuum had a 60 ms rise time and 110 ms formant
56

57
58 ³ Because of the limitations of the synthesizer, the step size was always a multiple of 10 ms.
59
60

1
2
3 transition (typical of the approximants, /l/ or /w/, as shown in Figure 2, left panels). The other
4
5 end of the continuum had a 10 ms rise time and a 40 ms formant transition (typical of the stops
6
7 /b/ or /d/, as shown in Figure 2, right panels). Intermediate values (Table 3) were chosen to span
8
9 the range between the approximant and stop “endpoints”.

10
11
12 Each stimulus had a steady-state vowel portion of the syllable that remained fixed as the
13
14 vowel /a/ with the following formant frequency (F) and bandwidth (b) values: F1 700 Hz; F2
15
16 1200 Hz; F3 2600 Hz; b1 130; b2 70; b3 160. The fundamental frequency (f_0) started at 120 Hz
17
18 and fell linearly to 80 Hz across the 300 ms syllable. The amplitude of voicing (av) was held
19
20 constant until it fell from 60 to 0 dB over the final 60 ms of the syllable.
21
22
23

24
25 The combination of five steps on spectral and five on the temporal dimension created 25
26
27 total stimuli. The combination of the most extreme values on each continuum created four
28
29 “endpoints” with spectral and temporal characteristics typical of /ba, da, wa, la/ (Figure 2).
30
31 Taken together, the two sets of stimuli sampled ability to use both lower- and higher-frequency
32
33 spectral cues as well as both static and dynamic spectral cues.
34
35

36 ***Test procedure***

37
38 Testing was conducted in a double-walled sound-treated booth. The stimuli were
39
40 converted from digital to analog format (TDT RX6) and passed through a programmable
41
42 attenuator (TDT PA5) and headphone buffer (TDT HB7) for presentation via an ER-2 insert
43
44 earphone to the participant’s right ear. Participants used a computer mouse to select which
45
46 endpoint they heard from a graphical display of four alternatives, labeled orthogonally as either
47
48 “see”, “shee”, “chee”, “tsee” (for stimulus set 1) or “dah”, “lah”, “bah”, “wah” (for stimulus set
49
50
51
52
53
54
55
56
57
58
59
60 2).

1
2
3 Stimulus set 1 was always tested first, followed by stimulus set 2. Within each stimulus
4 set the procedure consisted of three phases: familiarization, training and testing. Familiarization
5 was included to acquaint the listeners with the synthetic stimuli. During familiarization,
6 participants listened to a block of 40 trials where each trial was randomly chosen from the set of
7 four endpoint stimuli (10 each). The appropriate label was shown with green highlighting. The
8 participant was asked to confirm they could pair the stimulus and label by clicking on the
9 highlighted label. Following familiarization, the participant completed a training block of 40
10 trials. Each trial was randomly chosen from the set of four endpoint stimuli (10 each). After
11 each trial, the participant was asked to identify each token from the set of possible endpoint
12 labels. Visual correct answer feedback was provided after each response. Training continued
13 until 160 trials (4 blocks) had been completed or until the participant could achieve 80% correct
14 for each endpoint, whichever occurred first. For Experiment 1, only listeners with normal
15 hearing who achieved 80% correct for each endpoint (n=20 for stimulus set 1, n=16 for stimulus
16 set 2) proceeded to the test phase. During testing, participants completed 375 randomly selected
17 trials (15 trials of each token) without correct answer feedback. Rest breaks were given after 125
18 and 250 trials, and between the two stimulus sets.

Results

19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Figure 3 shows raw data for stimulus set 1. Recall that each token had two dimensions:
spectral (bandwidth and location of spectral energy) and temporal (rise time / duration), each
represented by one of five steps along a continuum. The x and y axes represent the five steps of
these continua. For ease of viewing each of the four response alternatives is shown in a separate
panel. Each point within a panel represents a single response. More points plotted for a single

1
2
3 spectral x temporal stimulus indicates that choice was made more frequently. A single point
4
5 indicates that choice was rarely made.
6
7

8 To interpret these data, consider how the frequency of a response changes as the spectral
9 and/or temporal dimensions are varied. For example, by comparing the top right and lower left
10 panels, we can see that the number of /f/ responses increases (and the number of /s/ responses
11 decreases) as the spectral peak and bandwidth are shifted to higher frequencies. The listeners
12 with normal hearing usually made clear distinctions between the stimuli, with high probability of
13 identifying a given endpoint and a clear demarcation at roughly the midpoint on each continuum.
14 For example, for a 2 ms rise time (step 5 on the temporal continua), steps 1 and 2 on the spectral
15 were identified as “tsee” almost all of the time; and steps 3-5 on the spectral continua were
16 identified as “chee” almost all of the time. The clear demarcation for the group also implies
17 similar response patterns across individuals.
18
19
20
21
22
23
24
25
26
27
28
29
30

31 We were most interested in the relative contributions of spectral and temporal cues to
32 consonant identification. Because both the independent and dependent measures were
33 categorical, a nonparametric analysis was appropriate. As a first step in that process, a bivariate
34 kernel density estimation (KDE) was generated⁴. Briefly, the KDE was used to estimate the
35 underlying probability density function of observed data. It was analogous to a two-dimensional
36 histogram but without the constraints introduced by specifying histogram bin width and origin.
37 The contribution of each data point in the two-dimensional (spectral x temporal) space was
38 translated into an area around that point, and all point-areas were integrated to obtain a final two-
39 dimensional distribution (Silverman, 1986). The model used a standard bivariate normal density
40 kernel and bandwidths were determined using the normal reference distribution function in R,
41 based on Venables and Ripley (2002). The outcome of the KDE for stimulus set 1 is shown in
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56

57
58 ⁴ ggplot2 package (Wickham, 2009) for R (2013).
59
60

1
2
3 Figure 4. Each contour line indicated approximately a 10% increase in the probability that a
4 given token will be judged as the token. The size of each contour was determined by variability;
5 larger contours and more space between contours indicated more variability, while smaller, more
6 tightly-spaced contours indicated less variability. Contiguous distribution contours indicated a
7 higher probability of confusion along the connected dimension. Figure 5 shows such data for
8 stimulus set 2.
9
10
11
12
13
14
15
16

17
18 From Figures 4 and 5, we can infer that (1) listeners with normal hearing were able to
19 distinguish the various stimuli in both the spectral and temporal domains, with a high degree of
20 certainty; (2) for both stimulus sets, listeners tended to make more distinct differentiations in the
21 spectral than in the temporal domain, particularly for stimulus set 2.
22
23
24
25
26

27 Although the KDE offered a convenient way to visualize and qualitatively examine the
28 data, it did not allow us to compare the relative weighing of cues in a quantitative way. In order
29 to quantitatively compare the use of cues, we undertook a separate analysis whereby the data
30 were submitted to a discriminant function analysis (DFA) (Lachenbruch & Goldstein, 1979). The
31 output of a DFA is an equation or set of equations which indicates to what degree each of the
32 input variables contributes to the classification and how robust that classification is. Because the
33 classification results of linear discriminant analysis (LDA) and quadratic discriminant analysis
34 were almost identical for these data (correlation=.99 for each of the stimulus sets), the simpler
35 LDA model was selected. Separation between judgments was tested to ensure that it was large
36 enough that a discriminant analysis was warranted. The classifications were significantly distinct
37 for an LDA, with Wilks' lambda of 0.149 ($p < .0001$).
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 LDA automatically determined an optimal combination of variables so that the first
54 equation (LD1) provided the greatest discrimination between groups (in this case, between the
55
56
57
58
59
60

1
2
3 four stimulus choices), and the second (LD2) provided the second greatest discrimination among
4
5 the groups. The proportion of trace for each equation is the percentage of data accounted for by
6
7 the given equation. Within each equation, the absolute value of the coefficient, or “weight”, can
8
9 be interpreted as reflecting the contribution of that variable (Hair, Black, Babin, & Anderson,
10
11 2010). Table 4 shows a summary of the results of the LDA.
12
13

14
15 *Stimulus set 1.* LD1 correctly classified 74% of the data. LD1 indicated that spectral
16
17 information was more influential than temporal information. The second equation, LD2,
18
19 classified the remaining 26% of the data. LD2 indicated that temporal information was more
20
21 influential than spectral information. When the LDA was calculated on an individual basis, LD1
22
23 accounted for 61-86% of the variance across participants. Spectral cue coefficients ranged from
24
25 1.2-1.4, and temporal cue coefficients ranged from 0.1-0.6. That is, each individual listener with
26
27 normal hearing showed a similar pattern, with the classification determined to a greater extent by
28
29 spectral cues. As a whole the analysis suggests that spectral cues were more important for
30
31 classification than temporal cues, but that they both played some role.
32
33
34

35
36 *Stimulus set 2.* For stimulus set 2, LD1 accounted for 63% of the data and indicated that
37
38 listeners relied more heavily on spectral information. LD2 accounted for 37% of the data, with
39
40 greater reliance on temporal information; albeit with more variability in cue use than for stimulus
41
42 set 1. When the LDA was calculated on an individual basis, LD1 accounted for 59%-87% of the
43
44 data. Spectral cue coefficients ranged from 0.1-1.2, and temporal cue coefficients ranged from
45
46 0.0-1.1. Individual data will be discussed later in this paper.
47
48
49

50 Discussion

51
52
53 Results indicated that the listeners with normal hearing were able to use both spectral and
54
55 temporal cues to classify stimuli, with a preference for spectral cues. That was the case whether
56
57
58
59
60

1
2
3 spectral cues were relatively static and broad-band (stimulus set 1) or more dynamic (stimulus
4
5 set 2). Moreover, this pattern was typical of all individuals in the group. This seems reasonable
6
7 considering that all listeners had normal hearing, with presumably normal frequency resolution.
8
9 The pattern might be quite different for listeners with hearing loss whose spectral resolution (and
10
11 perhaps their relative use of spectral and temporal cues) may not be the same. That population
12
13 was examined in Experiment 2.
14
15

16 17 18 **Experiment 2**

19
20 Results of Experiment 1 indicated that listeners with normal hearing were sensitive to
21
22 differences in both spectral and temporal dimensions. In this experiment, we were interested in
23
24 individuals with hearing loss; specifically, whether they demonstrated different patterns of cue
25
26 use than listeners with normal hearing, as well as whether there was greater variability among
27
28 individuals than had been the case for the normal-hearing group. Audibility of speech cues
29
30 presents an additional concern for this group. Accordingly, the design also included measures
31
32 of signal audibility.
33
34

35 36 37 **Method**

38 39 ***Participants***

40
41 Participants in the group with hearing impairment included 10 adults (4 female) with
42
43 mild to moderately-severe sensorineural hearing loss, ages 63-83 years (mean age 71 years). All
44
45 reported gradual, late-onset hearing loss typical of presbycusis. Eight of the participants had
46
47 symmetrical loss (defined as between-ear difference ≤ 10 dB at three consecutive frequencies),
48
49 and one ear was randomly selected for testing. Two participants had an asymmetrical loss and
50
51 for those two individuals, the better ear was selected for testing. The participants were a
52
53 representative sample of adults with hearing loss, with hearing thresholds that were typical for
54
55
56
57
58
59
60

1
2
3 their age (Cruickshanks et al., 1998) at octave frequencies from .5-4 kHz. Audiograms for the
4
5 test ears of the ten listeners with hearing loss are shown in Figure 6.
6
7

8 ***Stimuli***

10 Test signals were the stimulus set 1 and stimulus set 2 tokens described for Experiment 1.
11
12 To compensate for audibility loss due to high-frequency threshold elevation, a flat individual
13
14 gain was applied for each listener. Frequency shaping was deliberately not used to maintain the
15
16 spectral and temporal properties of the signals as designed. The individual gain was based upon
17
18 the spectral content of the stimuli, with the goal of achieving audible bandwidth through 5 kHz.
19
20 Because 5 kHz is not tested in a routine audiogram, we used a custom program to measure the
21
22 participants' 5000 Hz thresholds. This program utilized a 2-down/1-up adaptive tracking
23
24 procedure (Levitt, 1971) to estimate the level in dB SPL at which the subject could detect 71% of
25
26 the target tones of a given frequency. A computer monitor displayed a panel with two buttons,
27
28 and on each trial, listeners saw the two buttons marked "A" and "B". Each button illuminated
29
30 for 400 ms with a 500 ms interstimulus interval. The listener then chose which button had been
31
32 illuminated during the period in which a tone was presented, and used a computer mouse to click
33
34 on the button corresponding to that choice. Correct responses led to decreases in level, and two
35
36 incorrect responses in a row led to increases in level. A change in direction (an increase followed
37
38 by a decrease or a decrease followed by an increase) was termed a "reversal". For the first four
39
40 reversals, changes in level were 5 dB. After the fourth reversal, the changes were to 2 dB, and
41
42 eight more reversals were obtained. The levels at which these final eight reversals occurred were
43
44 averaged, and that average value was used as the estimated 5 kHz threshold. In pilot testing, we
45
46 verified that threshold values obtained with the custom program were very similar to those
47
48 obtained via conventional audiometry.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 After the 5 kHz threshold was obtained, presentation level was adjusted for each listener.
4
5 Presentation levels of the speech tokens ranged from 68 dB SPL to 92 dB SPL across the 10
6
7 listeners with hearing loss (Figure 7). Nine of the ten listeners reported that the initially-selected
8
9 presentation level was comfortable for them. One listener's presentation level was reduced 5 dB
10
11 for comfort. The same presentation level was used for both stimulus sets.
12
13
14

15 ***Test procedure***

16
17 Test setup was similar to that described for the listeners with normal hearing. Stimulus
18
19 set 1 was always tested first, followed by stimulus set 2. As in Experiment 1, each stimulus set
20
21 included three phases: familiarization, training and testing. In Experiment 1, 80% correct
22
23 endpoint identification was taken to confirm appropriate labeling and required to move forward
24
25 to the test phase. Here, we assumed that some of the listeners with hearing loss would be unable
26
27 to utilize some cues even after training. Accordingly, all participants proceeded to testing, either
28
29 because they had achieved 80% endpoint identification or had completed four training blocks.
30
31 Due to his limited availability, one of the 10 participants did not complete stimulus set 2. Final
32
33 scores were based on 375 randomly selected trials (15 trials of each token) without correct
34
35 answer feedback.
36
37
38
39
40

41 **Results**

42
43 As in Experiment 1, the data were first examined qualitatively using a bivariate kernel
44
45 density estimation (KDE). For stimulus set 1 (Figure 8), patterns were similar to those for
46
47 listeners with normal hearing, albeit with more variation (i.e. greater spread of the contours). The
48
49 strong vertical pattern suggests that for a given spectral token value, participants with hearing
50
51 loss were likely to classify it similarly, with greater variation in their classification based on the
52
53 temporal dimension. For stimulus set 2 (Figure 9), patterns were notably different than those for
54
55
56
57
58
59
60

1
2
3 the listeners with normal hearing (Figure 5). Figure 9 shows a large amount of variation as well
4 as a very different use of cues. The strong horizontal lines suggest that participants with hearing
5 impairment are more likely to agree about a token's temporal dimension (i.e., manner) than the
6 spectral dimension (i.e., place).
7
8
9

10
11
12 As in Experiment 1, we also examined the data quantitatively using a linear discriminant
13 analysis. The participants' classifications were significantly distinct for an LDA, with Wilks'
14 lambda of 0.194 ($p < .001$). Results of the analysis are summarized in Table 5.
15
16
17

18
19 For stimulus set 1, LD1 explained 68% of the variance. Spectral cues accounted for the
20 most variance (coefficient 1.17 compared to 0.16 for temporal cues). LD2 explained the
21 remaining 32% of the variance, with slightly greater emphasis on temporal cues. Overall the
22 analysis suggests a spectrally-mediated change, similar to the balance for the listeners with
23 normal hearing.
24
25
26
27
28
29
30

31
32 For stimulus set 2, LD1 explained 81% of the variance. Temporal cues accounted for the
33 most variance (coefficient 1.02 compared to 0.15 for spectral cues). LD2 explained a small
34 amount of variance (19%) with modestly greater reliance on spectral cues (coefficient 0.79
35 compared to -0.11 for temporal cues).
36
37
38
39

40
41 As was done in Experiment 1, we also calculated the LDA on an individual basis.
42 Results for LD1 (which explained the highest proportion of variance) are shown in Figure 10
43 (triangles), along with Experiment 1 data for the participants with normal hearing (circles). The
44 data are plotted as the absolute value of the coefficient or "weight" for temporal versus spectral
45 information. Participants for whom spectral and temporal cues were equally weighted would fall
46 on or very close to the diagonal reference line in each plot. Symbols falling above the reference
47 line indicated that categorization was based to a greater extent on temporal cues and symbols
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 falling below the reference line indicate a greater emphasis on spectral cues. For stimulus set 1,
4
5 there was a clear reliance on spectral cues for all of the listeners with normal hearing and the
6
7 majority of listeners with hearing loss, with both groups showing similar patterns ($\chi^2=2.07$,
8
9 $p=0.15$). However, there was also more variability across individuals than had been the case for
10
11 the listeners with normal hearing, with two listeners showing either greater use of temporal cues;
12
13 or nearly-equivalent use of both cues.
14
15

16
17 For stimulus set 2, there was much more variability in cue use. The majority of the
18
19 listeners with normal hearing continued to use spectral cues. Nearly all of the listeners with
20
21 hearing loss were more influenced by temporal cues, and the pattern of cue use was significantly
22
23 different than for the listeners with normal hearing ($\chi^2=6.51$, $p=0.01$). Only two listeners (#270
24
25 and #272) showed a normal-hearing like reliance on spectral cues for this stimulus set. Among
26
27 the listeners with hearing loss, those two participants had relatively better thresholds, particularly
28
29 in the high frequencies. They did not have greater signal audibility (in terms of the presentation
30
31 level of the signals re: their thresholds). However, their ability to use spectral cues to a greater
32
33 extent could reflect better spectral resolution due to narrower auditory filters.
34
35
36
37

38 Discussion

39
40 These data suggested that when audibility issues were addressed, the listeners with
41
42 hearing loss could use the static broad-band spectral cues even in the higher frequencies, but
43
44 were less able to use the dynamic spectral cues. Reduced use of dynamic spectral cues occurred
45
46 even when those cues were in the lower frequencies where the listeners had better thresholds
47
48 (and, perhaps, better resolution). The data also suggest that responses for listeners with hearing
49
50 loss were based to a relatively greater extent on temporal rather than spectral information
51
52
53 Finally, there was considerable variability among individuals.
54
55
56
57
58
59
60

General Discussion

This study was motivated, in part, by the expectation that individuals with hearing loss would differ in the extent to which they could resolve specific cues to speech. If that proved to be true, a related issue (not tested in this study) was our interest in a more customized approach to amplification parameters that would take individual differences into account. As a first step in this line of work we created a set of test stimuli that varied in spectral and temporal information. We hypothesized that (1) listeners with normal hearing would show high sensitivity to both dimensions, and those individuals would behave similarly to one another in this respect; and (2) that listeners with hearing impairment (who we assumed would have broadened auditory filters) would show reduced sensitivity to the spectral dimension, although with varying patterns across listeners.

Experiment 1 data supported the first hypothesis. Listeners with normal hearing were able to make distinctions across both temporal and spectral dimensions, although with a greater contribution from spectral cues. The second hypothesis was partially supported by Experiment 2. For stimulus set 1 where spectral cues were the frequency location of the frication noise, and temporal cues were envelope rise time, listeners with hearing loss were able to use both spectral and temporal cues but with a greater contribution from spectral cues (similar to the normal-hearing listeners). For stimulus set 2 where spectral cues were dynamic in nature (i.e., formant transitions) and temporal cues were envelope rise time, listeners with hearing loss relied more heavily on temporal cues. The data suggest that listeners with hearing loss showed greater weighting of the spectral dimension when the stimuli had static spectral cues, but a much greater use of the temporal cues when the stimuli had dynamic spectral cues.

1
2
3 Note that although when we varied the spectral and temporal dimensions independently
4 there was a relative effect which indicated sensitivity to change in the dimension being
5 manipulated, that effect may also indicate sensitivity to the way in which that dimension being
6 manipulated interacted with the other dimension. In other words, the response to a particular
7 stimulus must be determined by the total percept of both dimensions to the listener.

8 Accordingly, the difference between groups may reflect different reliance on temporal cues;
9 different reliance on dynamic cues; or even frequency-dependent differences.

10 ***Issues related to test stimuli***

11 We approached this project by devising a controlled test of speech sound identification,
12 in which specific temporal and spectral dimensions were varied. Synthetic speech was used to
13 provide precise control over the stimuli and to avoid two issues that may have been present in
14 earlier work: either confounding effects of accompanying cues, other than the cues under study
15 (e.g., Nitttrouer & Studdert-Kennedy, 1986); or artificial manipulation of cues in natural speech.

16 Although using manipulated natural speech is a valid approach to the problem, the mere act of
17 manipulating natural speech cues can introduce differences across subjects. For example, if
18 stimuli are adjusted for audibility by increasing the level of a specific cue (such as the consonant
19 burst), listeners may be influenced by that higher level to put more weight on that cue than would
20 be the case in a more natural-listening scenario. Accordingly, we attempted to put all listeners
21 on equal footing by creating new signals presented in similar ways and with equivalent training
22 to all participants.

23 With that said, synthetic speech may have its own effects. At least one study suggested
24 that older listeners may require more time to acclimatize to synthetic signals than younger
25 listeners (e.g., Wang & Humes, 2008). That study deliberately created non-speech-like stimuli
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 whose acoustic properties could not be interpreted as being speech sounds, so may not apply
4 directly to the present data. In addition, age-related differences in synthetic speech recognition
5
6 have been attributed to the higher thresholds that accompany aging, not to aging *per se* (Roring,
7
8 Hines, & Charness, 2007). The implication is that when audibility is controlled to compensate
9
10 for threshold elevation (as was the case here), older and younger listeners will react similarly to
11
12 synthetic speech materials. During training, the majority of our listeners in both experiments
13
14 reached criterion performance for the endpoint stimuli after similar numbers of trials, suggesting
15
16 that there were no overarching differences across groups. Nonetheless, it is possible that ability
17
18 to acclimatize to synthetic sounds varied across groups or individuals.
19
20
21
22
23

24
25 An important factor in speech-cue studies is audibility. Naturally, if a cue is inaudible to
26
27 a listener, it cannot be used. Audibility confounds—particularly for high-frequency cues—may
28
29 have been an unintended part of previous datasets (e.g., Coez et al., 2010). In support of the role
30
31 of audibility, when amplification is provided to listeners with hearing loss it improves use of
32
33 acoustic cues to a level similar to listeners with normal hearing (Calundruccio & Doherty, 2008;
34
35 Harkrider, Plyler, & Hedrick, 2006). In this case, we used the 5 kHz threshold (adaptive
36
37 threshold in a forced-choice paradigm) as a benchmark for audibility. Under those
38
39 circumstances, use of static spectral cues (stimulus set 1) by listeners with hearing loss was very
40
41 similar to listeners with normal hearing. That finding is in conflict with some earlier studies
42
43 which suggested such cues would not be well-used in cases of hearing loss, and suggest that
44
45 reduced audibility may have played a role in previous datasets.
46
47
48
49

50 51 ***Normal hearing and hearing loss***

52
53 The use of dynamic spectral cues (the formant transitions in stimulus set 2) showed some
54
55 differences across the two groups. Listeners with normal hearing used the formant transition to a
56
57
58
59
60

1
2
3 greater extent than envelope cue (as shown by the vertical lines in Figure 5; i.e., greater
4
5 probability of making a spectral than a temporal differentiation). Those data are in good
6
7 agreement with earlier work showing that normal-hearing listeners have full capability of using
8
9 formant transitions (Carpenter & Shahin, 2013; Hedrick & Younger, 2001, 2007; Lindholm et
10
11 al., 1988). When both cues were present for listeners with normal hearing, perceptions were
12
13 impacted by formant transition information to a greater extent than temporal (envelope) cues
14
15 (e.g., Nittrouer, Lowenstein, & Tarr, 2013). The difference between listeners with hearing loss
16
17 and normal hearing was dramatic; compare Figures 5 and 9, which show the heavy use of
18
19 temporal cues by listeners with hearing impairment. In a sense, listening to a definite formant
20
21 transition while having hearing loss may be a bit like a listener with normal hearing presented
22
23 with a neutral formant transition; that information is no longer available, necessitating other cue
24
25 use. However, our data were also in agreement with work showing that in the presence of
26
27 neutral formant transition information, listeners with normal hearing were able to shift to use of
28
29 temporal envelope information (Hedrick & Younger, 2001). Note that recent work by Nittrouer
30
31 and colleagues (2013) drew a different conclusion, in that their adult listeners with normal
32
33 hearing did not use envelope cues at all. That finding can be explained by considering
34
35 Nittrouer's stimuli, in which the formant transition was either typical of /ba/ or typical of /wa/
36
37 but never neutral. Although unlikely to arise in everyday speech, we can infer that in
38
39 conflicting-cue situation listeners with normal hearing appear to show a preference for use of
40
41 spectral rather than temporal information.
42
43
44
45
46
47
48
49

50 All listeners showed sensitivity to temporal cues, even when those were used as a
51
52 secondary cue. This is in good agreement with studies that hearing loss produces greater spectral
53
54 than temporal degradations, at least when the available temporal cues are contained in the speech
55
56
57
58
59
60

1
2
3 envelope (e.g., Turner, Souza, & Forget, 1995). One caveat is that the same principles may not
4
5 apply to other types of temporal cues, such as gaps, which may be affected by hearing loss
6
7 and/or aging to a greater extent (Walton, 2010).
8
9

10 ***Individual patterns***

11
12 Among listeners with normal hearing the relative use of spectral vs temporal cues was
13
14 quite similar. As expected, individuals with hearing loss showed more listener-to-listener
15
16 variation. The most interesting effects occurred among listeners with hearing loss presented with
17
18 stimulus set 2. Previous studies which addressed only group effects suggest that use of dynamic
19
20 spectral information might be impaired in all listeners. Instead, two of our hearing-impaired
21
22 listeners showed responses much like the listeners with normal hearing. Those individuals had
23
24 similar age, etiology and duration of loss as the rest of the hearing-impaired group but had
25
26 relatively better thresholds, particularly in the high frequencies. However, they did not have
27
28 greater signal audibility (in terms of the presentation level of the signals re: their thresholds).
29
30 Perhaps their ability to use spectral cues to a greater extent reflected better spectral resolution
31
32 due to narrower auditory filters, or some other auditory (or listening strategy) difference.
33
34
35
36
37

38
39 We recruited a representative sample of adults with hearing loss, who, statistically, were
40
41 likely to be older. Because this study was not designed to address contributions of age versus
42
43 hearing loss, the data leave unresolved the question of whether an older listener with a
44
45 completely normal audiogram would perform like the Experiment 1 subjects, or show variability
46
47 of specific cue use more like the Experiment 2 subjects. This is an interesting question and
48
49 depends somewhat on one's view. If we take the position that greater dependence on temporal
50
51 cues is a direct consequence of poor spectral resolution, we might expect older listeners with
52
53 normal hearing, who show little evidence of auditory filter broadening (Ison, Virag, Allen, &
54
55
56
57
58
59
60

1
2
3 Hammond, 2002), to maintain good spectral resolution. On the other hand, some authors (e.g.
4
5 Pichora-Fuller, Schneider, Macdonald, Pass, & Brown, 2007) have suggested that asynchronous
6
7 nerve firing which accompanies aging can degrade representation of temporal fine structure,
8
9 which could impact use of the “spectral” cues presented here.
10
11

12 ***Implications for hearing rehabilitation***

13
14
15 The stimuli used in this study could potentially be used to identify listeners who are more
16
17 spectral-dependent versus listeners who are more temporal-dependent. Although this area is
18
19 largely unexplored, if we have the ability to measure listener cue “weighting” we could better
20
21 understand the consequences of that listener’s specific listening habits, and perhaps modify
22
23 therapeutic interventions. One consequence of modern amplification technology is a tendency to
24
25 distort in the spectral domain, the temporal domain, or both. Many technologies distort one
26
27 dimension to a greater extent. For example, fast-acting compression can degrade envelope cues.
28
29 A listener who depends heavily on temporal envelope vs spectral cues may be affected by even
30
31 minimal envelope distortion (Souza, Hoover, & Gallun, 2012). Conversely, a subject who bases
32
33 categorical perception largely on spectral cues may tolerate larger amounts of envelope distortion
34
35 without impaired recognition. Multichannel compression with a high number of compression
36
37 channels and a relatively high compression ratio has many advantages, such as more specific
38
39 noise management and better loudness equalization, but can smooth spectral contrasts and
40
41 degrade recognition when the essential information is carried by spectral contrasts and presented
42
43 to a listener who is sensitive to those contrasts (Souza, Wright, et al., 2012). It may even be
44
45 possible to shift cue dependence with training (Francis, Baldwin, & Nusbaum, 2000). Such
46
47 topics offer an interesting area for future translational work.
48
49
50
51
52
53
54

55 **Acknowledgments**

56
57
58
59
60

This work was supported by NIH R01 DC60014.

References

- ANSI. (2004). *American National Standards Institute Specification for Audiometers* (S3.6-2004).
- 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.
- 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.
- Boothroyd, A., Springer, N., Smith, L., & Schulman, J. (1988). Amplitude compression and profound hearing loss. *Journal of Speech and Hearing Research*, *31*, 362-376.
- Bor, S., Souza, P., & Wright, R. (2008). Multichannel compression: Effects of reduced spectral contrast on vowel identification. *Journal of Speech, Language and Hearing Research*, *51*, 1315-1327.
- Brennan, M. A., Gallun, F. J., Souza, P. E., & Stecker, G. C. (2013). Temporal resolution with a prescriptive fitting formula. *American Journal of Audiology*, *22*, 216-225.
- Calundruccio, L., & Doherty, K. A. (2008). Spectral weighting strategies for hearing-impaired listeners measures using a correlational method. *Journal of the Acoustical Society of America*, *123*, 2367-2378.

1
2
3
4
5
6
7
8
9
Carpenter, A. L., & Shahin, A. J. (2013). Development of the N1-P2 auditory evoked
response to amplitude rise time and rate of formant transition of speech sounds. *Neuroscience
Letters, 544*, 56-61.

10
11
12
13
14
15
16
Christensen, L. A., & Humes, L. E. (1997). Identification of multidimensional stimuli
containing speech cues and the effects of training. *Journal of the Acoustical Society of America,*
102, 2297-2310.

17
18
19
20
21
22
23
24
Coez, A., Belin, P., Bizaguet, E., Ferrary, E., Zibovicus, M., & Samson, Y. (2010).
Hearing loss severity: impaired processing of formant transition duration. *Neuropsychologica,*
48, 3057-3061.

25
26
27
28
29
30
31
32
33
Cruikshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E., Klein, R., Mares-Perlman,
J. A., & Nondahl, D. M. (1998). Prevalence of hearing loss in older adults in Beaver Dam,
Wisconsin. The epidemiology of hearing loss study. *American Journal of Epidemiology, 148*,
879-886.

34
35
36
37
38
39
40
Davies-Venn, E., & Souza, P. (2014). The role of spectral resolution, working memory,
and audibility in explaining variance in susceptibility to temporal envelope distortion. *Journal of
the American Academy of Audiology, 25*, 592-604.

41
42
43
44
45
46
47
Davies-Venn, E., Souza, P., Brennan, M., & Stecker, G. C. (2009). Effects of audibility
and multichannel wide dynamic range compression on consonant recognition for listeners with
severe hearing loss. *Ear and Hearing, 30*, 494-504.

48
49
50
51
52
53
54
55
56
57
58
59
60
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.

1
2
3 Dubno, J. R., Dirks, D. D., & Ellison, D. E. (1989). Stop-consonant recognition for
4 normal-hearing listeners and listeners with high-frequency hearing loss. I: the contribution of
5 selected frequency regions. *Journal of the Acoustical Society of America*, 85, 347-354.
6
7

8
9
10 Faulkner, A., Rosen, S., & Moore, B. C. (1990). Residual frequency selectivity in the
11 profoundly hearing-impaired listener. *British Journal of Audiology*, 24, 381-392.
12
13

14
15 Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to
16 acoustic cues. *Perception & Psychophysics*, 62, 1668-1680.
17
18

19
20 Glasberg, B. R., & Moore, B. C. (1986). Auditory filter shapes in subjects with unilateral
21 and bilateral cochlear impairments. *Journal of the Acoustical Society of America*, 79, 1020-1033.
22
23

24
25 Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for
26 units in speech perception. *Journal of Phonetics*, 31, 305-320.
27
28

29
30 Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data*
31 *analysis* (Seventh Edition). Upper Saddle River, NJ: Prentice Hall.
32
33

34
35 Harkrider, A. W., Plyler, P. N., & Hedrick, M. (2006). Effects of hearing loss and
36 spectral shaping on identification and neural response patterns of stop-consonant stimuli. *Journal*
37 *of the Acoustical Society of America*, 120, 915-925.
38
39

40
41 Hedrick, M. (1997). Effect of acoustic cues on labeling fricatives and affricates. *Journal*
42 *of Speech, Language and Hearing Research*, 40, 925-938.
43
44

45
46 Hedrick, M., & Rice, T. (2000). Effect of a single-channel wide dynamic range
47 compression circuit on perception of stop consonant place of articulation. *Journal of Speech,*
48 *Language and Hearing Research*, 43, 1174-1184.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Hedrick, M., & Younger, M. S. (2001). Perceptual weighting of relative amplitude and
4 formant transition cues in aided CV syllables. *Journal of Speech, Language and Hearing*
5
6
7
8
9 *Research, 44*, 964-974.

10 Hedrick, M., & Younger, M. S. (2003). Labeling of /s/ and /sh/ by listeners with normal
11 and impaired hearing, revisited. *Journal of Speech, Language and Hearing Research, 46*, 636-
12
13
14
15
16
17 648.

18 Hedrick, M., & Younger, M. S. (2007). Perceptual weighting of stop consonant cues by
19 normal and impaired listeners in reverberation versus noise. *Journal of Speech, Language and*
20
21
22
23
24
25
26
27
28 *Hearing Research, 50*, 254-269.

29 Howell, P., & Rosen, S. (1983). Production and perception of rise time in the voiceless
30 affricate/fricative distinction. *Journal of the Acoustical Society of America, 73*, 976-984.

31 Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit
32 provided by amplified speech to older adults. *Journal of the American Academy of Audiology,*
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
18, 590-603.

Ison, J. R., Virag, T. M., Allen, P. D., & Hammond, G. R. (2002). The attention filter for
tones in noise has the same shape and effective bandwidth in the elderly as it has in young
listeners. *Journal of the Acoustical Society of America, 112*, 238-246.

Jenstad, L. M., & Souza, P. E. (2005). Quantifying the effect of compression hearing aid
release time on speech acoustics and intelligibility. *Journal of Speech, Language and Hearing*
Research, 48, 651-667.

Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the*
Acoustical Society of America, 67, 971-995.

1
2
3 Lachenbruch, P. A., & Goldstein, M. (1979). Discriminant analysis. *Biometrics*, 35, 69-
4
5
6 85.

7
8 LaRiviere, C., Winitz, H., & Herriman, E. (1975). The distribution of perceptual cues in
9
10 English prevocalic fricatives. *Journal of Speech and Hearing Research*, 18, 613-622.

11
12 Leek, M. R., & Summers, V. (1996). Reduced frequency selectivity and the preservation
13
14 of spectral contrast in noise. *Journal of the Acoustical Society of America*, 100, 1796-1806.

15
16
17 Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the*
18
19 *Acoustical Society of America*, 49(2), 467.

20
21
22 Lindholm, J. M., Dorman, M. F., Taylor, B. E., & Hannley, M. T. (1988). Stimulus
23
24 factors influencing the identification of voiced stop consonants by normal-hearing and hearing-
25
26 impaired adults. *Journal of the Acoustical Society of America*, 83, 1608-1614.

27
28
29 Molis, M. R., & Leek, M. R. (2011). Vowel identification by listeners with hearing
30
31 impairment in response to variation in formant frequencies. *Journal of Speech, Language and*
32
33 *Hearing Research*, 54, 1211-1223.

34
35
36 Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical*
37
38 *Society of America*, 101, 3241-3254.

39
40
41 Nittrouer, S., Lowenstein, J. H., & Tarr, E. (2013). Amplitude rise time does not cue the
42
43 /ba/-/wa/ contrast for adults or children. *Journal of Speech, Language and Hearing Research*, 56,
44
45 427-440.

46
47
48 Nittrouer, S., & Studdert-Kennedy, M. (1986). The stop-glide distinction: acoustic
49
50 analysis and perceptual effect of variation in syllable amplitude envelope for initial /b/ and /w/.
51
52 *Journal of the Acoustical Society of America*, 80, 1026-1029.

53
54
55
56
57
58
59
60

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.

Preminger, J., & Wiley, T. L. (1985). Frequency selectivity and consonant intelligibility in sensorineural hearing loss. *Journal of Speech and Hearing Research*, 28, 197-206.

Reed, C. M., Braida, 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 suprathreshold deficits. *Trends in Amplification*, 13, 4-43.

Repp, B. H. (1983). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication*, 2, 341-361.

Roring, R. W., Hines, F. G., & Charness, N. (2007). Age differences in identifying words in synthetic speech. *Human Factors*, 49, 25-31.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *ICASSP 1997*, 336, 367-373.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Singh, R., & Allen, J. B. (2012). The influence of stop consonants' perceptual features on the Articulation Index model. *Journal of the Acoustical Society of America*, 131, 3051-3068.

Souza, P., Boike, K. T., Witherell, K., & Tremblay, K. (2007). Prediction of speech recognition from audibility in older listeners with hearing loss: effects of age, amplification, and background noise. *Journal of the American Academy of Audiology*, 18, 54-65.

1
2
3 Souza, P., Hoover, E., & Gallun, F. J. (2012). Application of the envelope difference
4 index to spectrally sparse speech. *Journal of Speech, Language and Hearing Research*, 55, 824-
5
6
7
8 837.

9
10 Souza, P., Jenstad, L. M., & Folino, R. (2005). Using multichannel wide-dynamic range
11 compression in severely hearing-impaired listeners: effects on speech recognition and quality.
12
13
14
15 *Ear and Hearing*, 26, 120-131.

16
17 Souza, P., & Kitch, V. (2001). The contribution of amplitude envelope cues to sentence
18 identification in young and aged listeners. *Ear and Hearing*, 22, 112-119.

19
20 Souza, P., Wright, R., & Bor, S. (2012). Consequences of broad auditory filters for
21 identification of multichannel-compressed vowels. *Journal of Speech, Language and Hearing*
22
23
24
25
26
27
28 *Research*, 55, 474-486.

29
30 Stelmachowicz, P. G., Kopun, J., Mace, A., Lewis, D. E., & Nittrouer, S. (1995). The
31 perception of amplified speech by listeners with hearing loss: acoustic correlates. *Journal of the*
32
33
34
35
36
37
38 *Acoustical Society of America*, 98, 1388-1399.

39
40 Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., & Kurowski, K. (1992).
41 Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of*
42
43
44
45
46
47
48 *the Acoustical Society of America*, 91, 2979-3000.

49
50 Team, R. C. (2013). R: *A language and environment for statistical computing*. Vienna,
51
52
53
54
55
56
57
58
59
60 Austria. Retrieved from <http://www.R-project.org>

Turner, C. W., & Henn, C. C. (1989). The relation between vowel recognition and
measures of frequency resolution. *Journal of Speech and Hearing Research*, 32, 49-58.

1
2
3 Turner, C. W., Smith, S. J., Aldridge, P. L., & Stewart, S. L. (1997). Formant transition
4 duration and speech recognition in normal and hearing-impaired listeners. *Journal of the*
5
6 *Acoustical Society of America*, 101, 2822-2825.
7
8

9
10 Turner, C. W., Souza, P., & Forget, L. N. (1995). Use of temporal envelope cues in
11
12 speech recognition by normal and hearing-impaired listeners. *Journal of the Acoustical Society*
13
14 *of America*, 97, 2568-2576.
15
16

17 Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics* (Fourth Edition).
18

19 Walton, J. P. (2010). Timing is everything: temporal processing deficits in the aged
20
21 auditory brainstem. *Hearing Research*, 264, 63-69.
22
23

24 Wang, X., & Humes, L. (2008). Classification and cue weighting of multidimensional
25
26 stimuli with speech-like cues for young normal hearing and elderly hearing-impaired listeners.
27
28 *Ear and Hearing*, 29, 725-745.
29
30

31 Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer.
32
33

34 Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R.
35
36 Kircher & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34-57). Cambridge: Cambridge
37
38 University Press.
39

40 Zeng, F. G., & Turner, C. W. (1990). Recognition of voiceless fricatives by normal and
41
42 hearing-impaired subjects. *Journal of Speech and Hearing Research*, 33, 440-449.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Formant (F), bandwidth (b), and formant-amplitude (a) values in the alveolar (/s/-/ts/) to palatal (/ʃ/-/tʃ/) spectral continuum

	alveolar				palatal
F1	320	315	310	305	300
F2	1396	1507	1618	1729	1840
F3	2530	2585	2640	2695	2750
b1	200	200	200	200	200
b2	80	84	88	92	100
b3	200	225	250	275	300
a3	0	14	28	42	56
a4	0	12	24	36	48
a5	0	12	24	36	48
a6	51	50	49	48	47

Table 2. Formant (F) and bandwidth (b) values for stimulus set 2

	alveolar				labial
F1	250	250	250	250	250
F2	1500	1338	1175	1013	850
F3	2700	2563	2425	2288	2150
F4	3250	3250	3250	3250	3250
F5	3700	3700	3700	3700	3700
b1	55	55	55	55	55
b2	100	99	98	96	95
b3	725	568	410	253	95
b4	200	200	200	200	200
b5	200	200	200	200	200

Table 3. Starting amplitude of voicing (av), rise-time duration to maximum voicing amplitude of 60 dB, and duration of formant transitions for stimulus set 2.

Manner	starting av value	rise time	formant transition
stop	60	10 ms	40 ms
	50	20 ms	50 ms
	40	30 ms	70 ms
	40	40 ms	90 ms
approximant	40	60 ms	110 ms

Table 4. Group results of Linear Discriminant Analysis for listeners with normal hearing

<i>Stimulus set 1</i>		LD1	LD2
Discriminant weight	Spectral	-1.32	-0.11
	Temporal	-0.14	0.97
Proportion of trace		0.74	0.26
<i>Stimulus set 2</i>		LD1	LD2
Discriminant weight	Spectral	0.96	0.34
	Temporal	0.39	-0.85
Proportion of trace		0.63	0.37

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 5. Results of Linear Discriminant Analysis for listeners with hearing impairment

<i>Stimulus set 1</i>		LD1	LD2
Discriminant weight	Spectral	1.17	0.13
	Temporal	0.16	-0.95
Proportion of trace		0.68	0.32
<i>Stimulus set 2</i>		LD1	LD2
Discriminant weight	Spectral	0.15	0.79
	Temporal	1.03	-0.11
Proportion of trace		0.81	0.19

Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure legends

Figure 1. Waveforms and spectrograms of the four stimulus endpoints with the temporal (fricative-affricate) dimension on the x-axis and the frequency (alveolar-palatal) dimension on the y-axis.

Figure 2. Waveforms and spectrograms of the four stimulus endpoints with the temporal (stop~approximant) dimension on the x axis and the frequency (alveolar~labial) dimension on the y axis. Note the difference in formant location and transition near the onset of each stimulus.

Figure 3. Stimulus set 1 responses for listeners with normal hearing. Each panel represents the 25 possible spectral x temporal continua steps. Each point represents a single response. More frequent responses to a category are plotted as larger groupings of points. For example, listeners tended to make more “see” responses when the spectrum contained lower-frequency energy and the envelope had a long rise time (lower left corner of the top right “See” panel). For ease of viewing, axes are labeled using one of the acoustic parameters; for full parameter set, see Tables 2-3.

Figure 4. Results of KDE analysis (2d density estimation function from the ggplot2 package [Wickham, 2009]) for listeners with normal hearing, for stimulus set 1. Each classification is shown in its own graph for ease of reading, but the scales and axes are the same. A point in the graph indicates that at least one judgment of that type was observed for that combination of temporal and spectral cues. Each contour line represents approximately a 10% increase in likelihood of the given judgment for each point it encompasses. For example, the probability of a of a normal-hearing listener judging a token with a spectral and temporal value of 2 as “see” is near 100%.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 5. Results of KDE analysis as in Figure 5, but for listeners with normal hearing for stimulus set 2. Note the contiguous distributions along the temporal dimension, suggesting a greater certainty when identifying signals based on spectral cues (formants and/or bandwidth). There was also greater variability in judgments, which can be seen in the increased size of the contours.

Figure 6. Audiograms for listeners with hearing impairment. Thin lines show individual audiograms; thick line/filled circles shows the mean audiogram for the group.

Figure 7. Illustration of speech audibility for two subjects, roughly representing the range of audiograms (least hearing loss to most hearing loss) among the study cohort. Each panel shows pure-tone thresholds (circles) for a single participant, relative to 1/3 octave band levels measured for /s/ at that participant's individually-set presentation level (squares). All levels are expressed as dB SPL measured in a 2cc coupler.

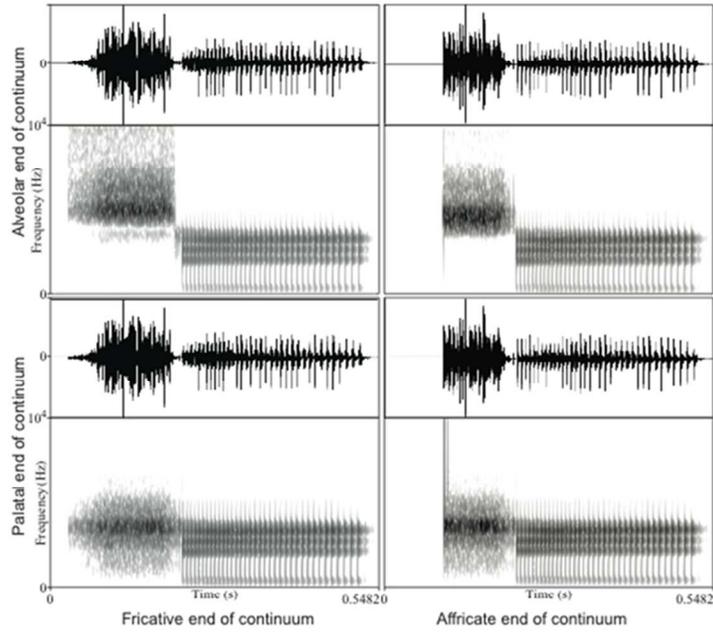
Figure 8. Results of KDE analysis (2d density estimation function from the ggplot2 package [Wickham 2009]) for listeners with hearing loss, for stimulus set 1. Each classification is shown in its own graph for ease of reading, but the scales and axes are the same. A point in the graph indicates that at least one judgment of that type was observed for that combination of temporal and spectral cues. Each contour line represents approximately a 10% increase in likelihood of the given judgment for each point it encompasses. Variability is shown by the size and spread of contours; small, tight contours indicate high certainty.

Figure 9. Results of KDE analysis as in Figure 8, but for listeners with hearing loss for stimulus set 2. Note the contiguous distributions along the temporal dimension (i.e. the contours cover a large number of tokens with little separation), suggesting difficulty using temporal cues to identify sounds.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 10. Individual spectral and temporal weights for listeners with normal hearing (open circles) and hearing loss (filled triangles). All values are plotted as the absolute value of the individual coefficients for LD1. Left panel shows results for stimulus set 1 and right panel for stimulus set 2. Data points falling along the diagonal line in each panel indicate equal consideration of both stimulus dimensions. Data points falling below and to the right of the diagonal indicate greater reliance on spectral cues. Data points falling above and to the left of the diagonal indicate greater reliance on temporal cues.

For Peer Review



Figure

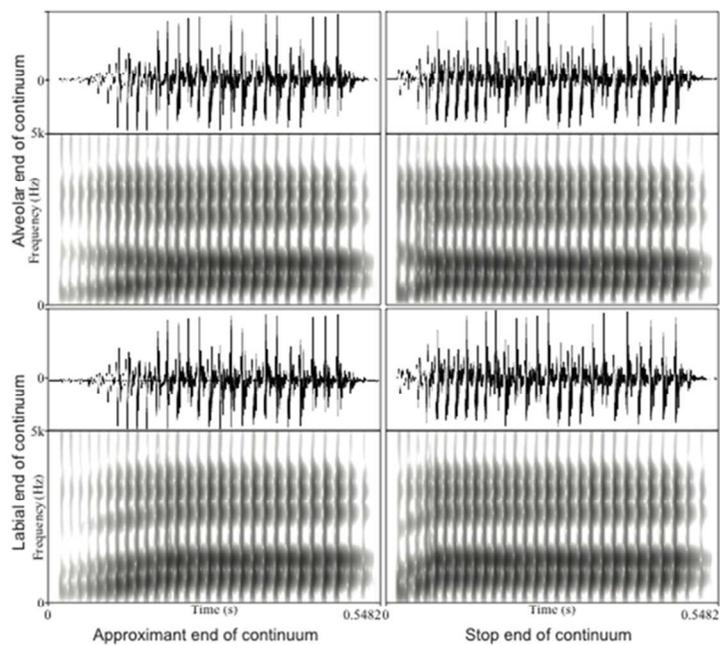
1

Fig 1
254x190mm (96 x 96 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

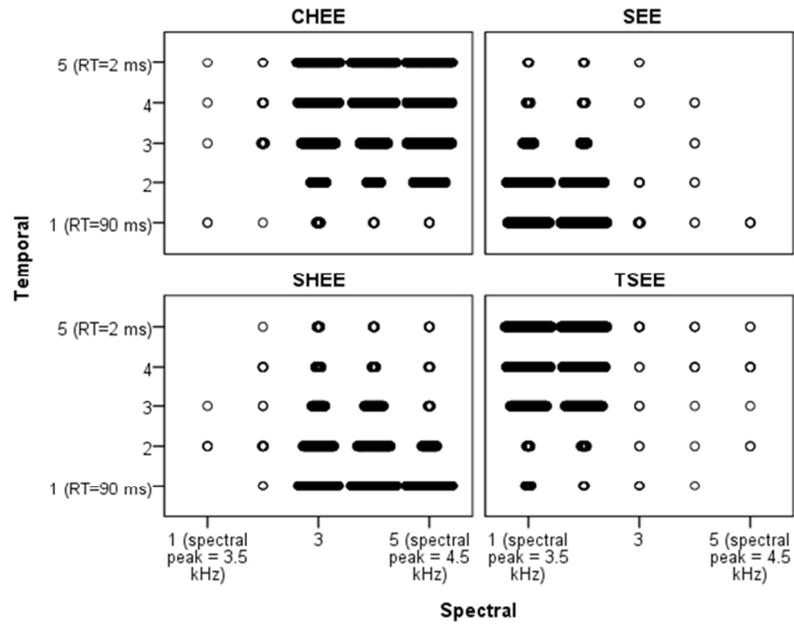


Figure

2

Fig 2
254x190mm (96 x 96 DPI)

Review



Figure

3

Fig 3
254x190mm (96 x 96 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

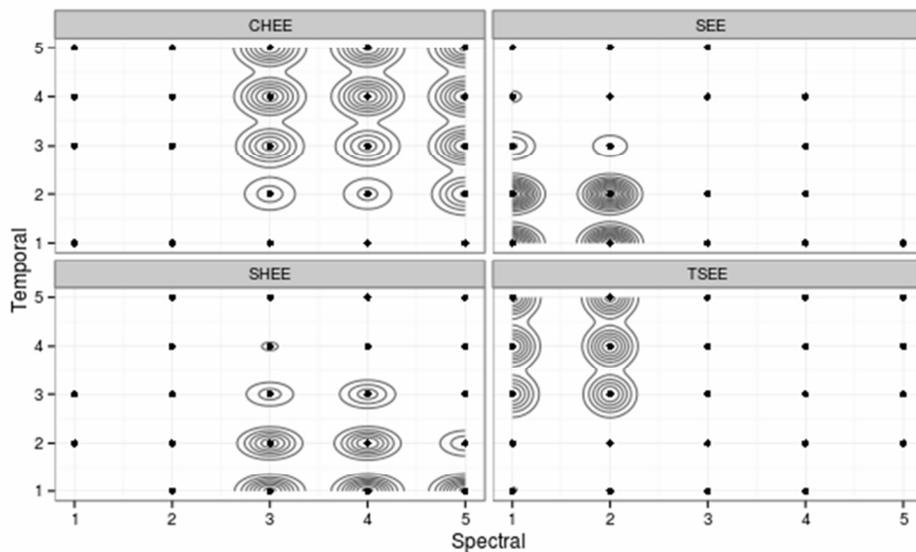


Figure 4

Fig 4
254x190mm (96 x 96 DPI)

Review

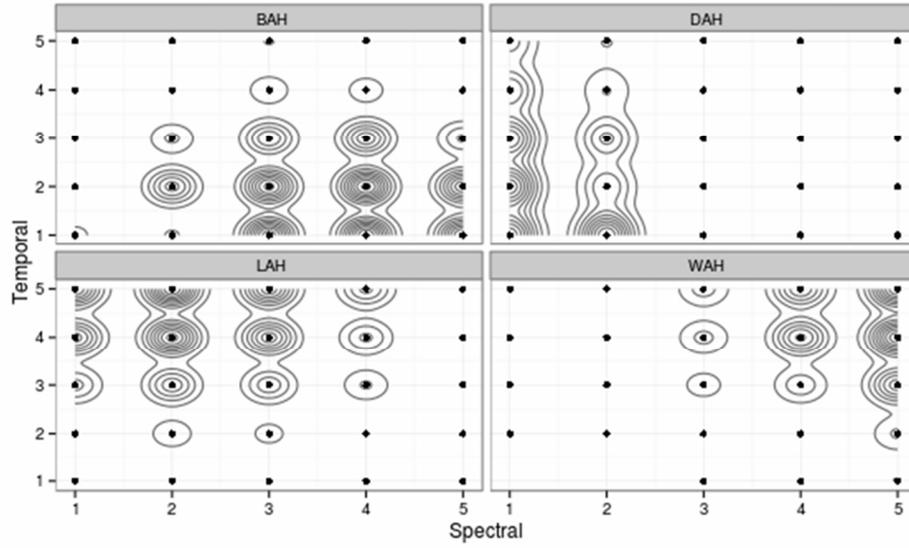


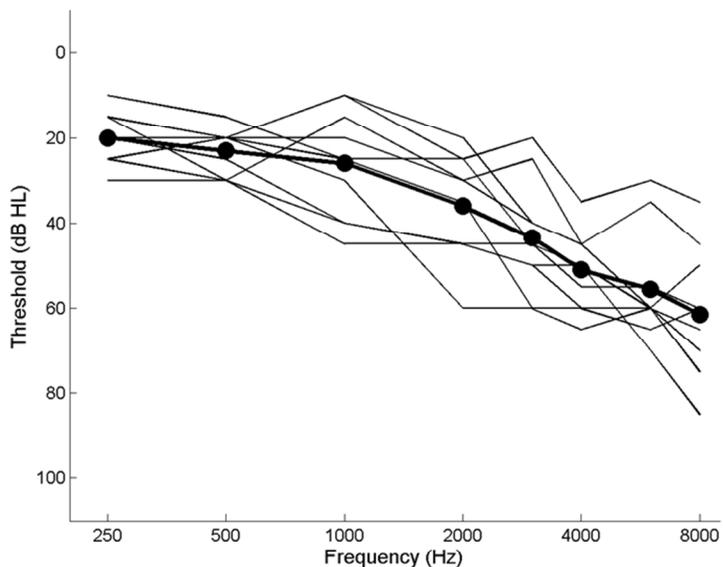
Figure 5

Fig 5
254x190mm (96 x 96 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Figure

6

Fig 6
254x190mm (96 x 96 DPI)

Review

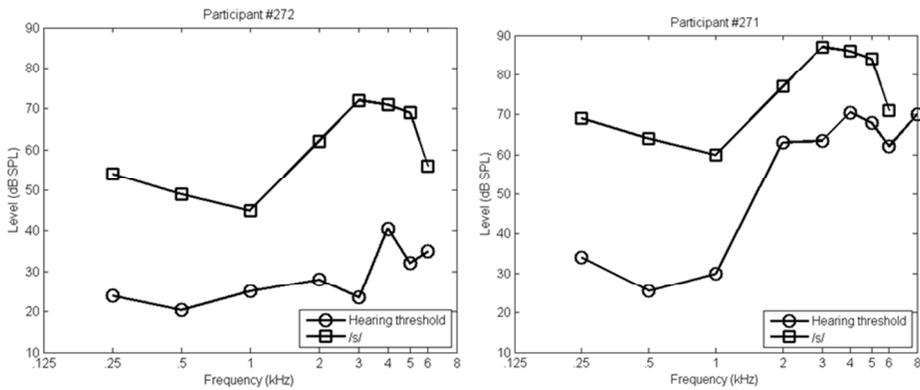


Figure 7

Fig 7
254x190mm (96 x 96 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

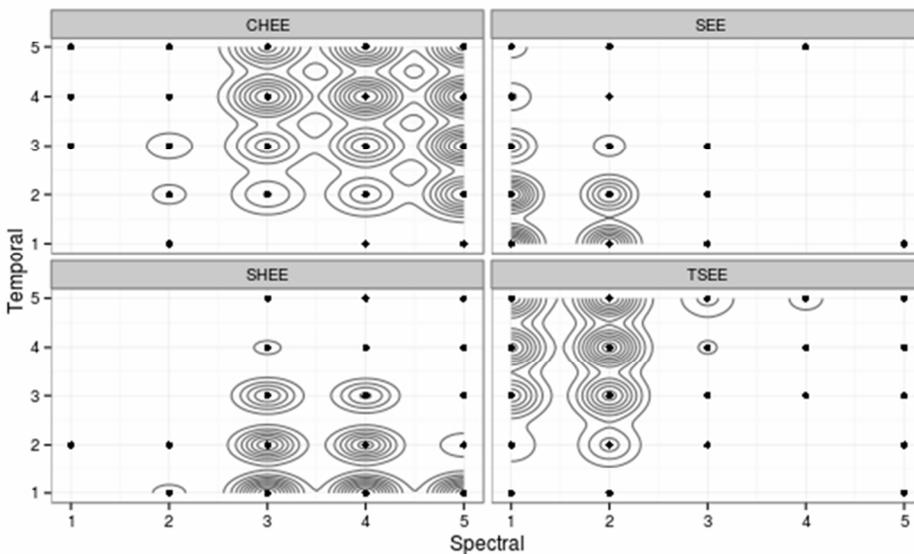


Figure 8

Fig 8
254x190mm (96 x 96 DPI)

Review

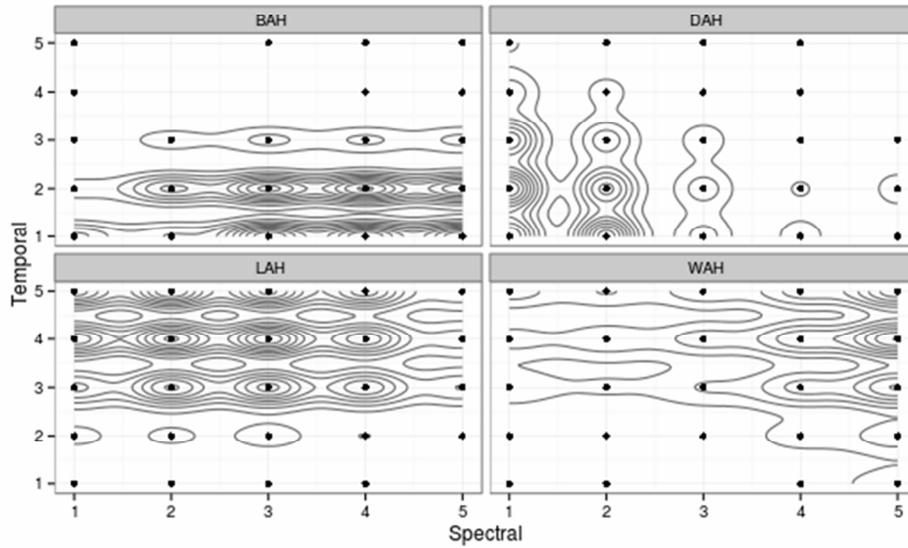


Figure 9

Fig 9
254x190mm (96 x 96 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

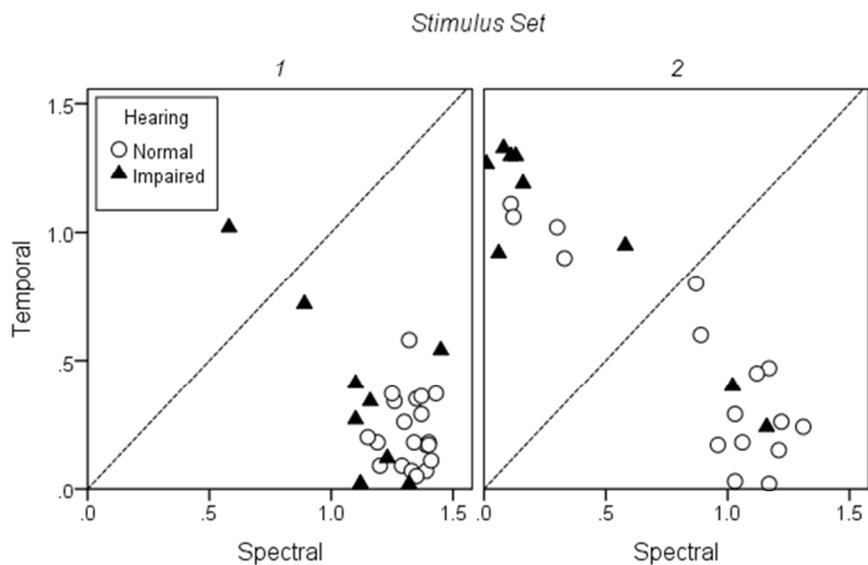


Figure 10

Fig 10
254x190mm (96 x 96 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60