Consonant categorization exhibits a graded influence of surrounding spectral context

Christian E. Stilp and Ashley A. Assgari

Citation: J. Acoust. Soc. Am. **141**, EL153 (2017); doi: 10.1121/1.4974769
View online: http://dx.doi.org/10.1121/1.4974769
View Table of Contents: http://asa.scitation.org/toc/jas/141/2
Published by the Acoustical Society of America

Consonant categorization exhibits a graded influence of surrounding spectral context

Christian E. Stilp[a)] and Ashley A. Assgari
Department of Psychological and Brain Sciences, University of Louisville, Louisville, Kentucky 40292, USA
christian.stilp@louisville.edu, ashley.assgari@louisville.edu

Abstract: When spectral properties differ across successive sounds, this difference is perceptually magnified, resulting in spectral contrast effects (SCEs). Recently, Stilp, Anderson, and Winn [(2015) J. Acoust. Soc. Am. **137**(6), 3466–3476] revealed that SCEs are graded: more prominent spectral peaks in preceding sounds produced larger SCEs (i.e., category boundary shifts) in categorization of subsequent vowels. Here, a similar relationship between spectral context and SCEs was replicated in categorization of voiced stop consonants. By generalizing this relationship across consonants and vowels, different spectral cues, and different frequency regions, acute and graded sensitivity to spectral context appears to be pervasive in speech perception.

[MC]
Date Received: November 10, 2016 **Date Accepted:** December 15, 2016

1. Introduction

All perception occurs in some form of context, and speech perception is no exception. Speech sounds are recognized by both their acoustic properties and how those relate to acoustic properties of the surrounding acoustic context (intrinsic and extrinsic cues, respectively; e.g., Ainsworth, 1975; Nearey, 1989). When spectral properties change between the preceding acoustic context and the subsequent target sound, these spectral differences are perceptually magnified, resulting in spectral contrast effects (SCEs).[1] For example, a preceding sentence with low-F_1 frequencies emphasized will produce more /ɛ/ (high F_1) responses to a target vowel, whereas a sentence with high-F_1 frequencies emphasized will produce more /ɪ/ (low F_1) responses to the target vowel (Ladefoged and Broadbent, 1957; Watkins, 1991; Sjerps *et al.*, 2011). In both cases, perception is biased away from stable spectral properties in the preceding sounds. SCEs influence categorization of a wide range of speech sound contrasts (see Stilp *et al.*, 2015 for review).

Historically, SCEs were largely considered to be binary phenomena: they either occurred (i.e., produced statistically significant shifts in speech sound categorization) or did not occur. Little to no consideration was given to the magnitude of the SCE itself, or how that magnitude might be predictable. Recently, Stilp *et al.* (2015) reported that SCE magnitudes in vowel categorization could be predicted from acoustic properties of the precursor sentence. The precursor sentence was filtered to feature a narrow spectral peak (100 Hz bandwidth), a broader spectral peak (300 Hz bandwidth), or a spectral profile reflecting the difference between spectral envelopes of /ɪ/ and /ɛ/ (i.e., spectral envelope difference filter; Watkins, 1991). Filter gain was also manipulated to vary the prominence of the added spectral properties (+5 to +20 dB for spectral peaks, 25% to 100% of the spectral envelope difference for difference filters). SCEs were highly linear across these manipulations, with effect magnitudes scaling with the size of the reliable spectral property in the precursor sentence. Vowel categorization was even influenced by very subtle manipulations of the precursor sentence (filtering with +5 dB gain or only 25% of the spectral envelope difference between vowel exemplars). From this evidence, Stilp *et al.* (2015) suggested that SCEs might be influencing speech perception more frequently than previously considered.

While these results are suggestive of a close coupling between spectral characteristics of a speech context and speech perception, questions of generalizability arise. First, the graded influence of spectral context was only explored in categorization of vowels. Vowel perception has long been considered to be less categorical than consonant perception (Fry *et al.*, 1962; Pisoni, 1973). As such, malleable category

[a)]Author to whom correspondence should be addressed.

boundaries for vowels might be more sensitive to the surrounding spectral context (producing graded SCEs), whereas sharper category boundaries for consonants might be less sensitive to such an influence (producing binary SCEs). Second, this fine-grained sensitivity to spectral context was only shown for a particular frequency region (below 1000 Hz), and for only one spectral cue (F_1). This draws into question whether SCE magnitudes can be predicted from signal acoustics more generally.

To address this question, we investigated sensitivity to spectral context during categorization of voiced stop consonants, /d/ and /g/. While prior research has shown that SCEs influence categorization of these consonants (Holt, 2005, 2006; Laing *et al.*, 2012), the present effort sought to characterize the nature of this influence and predict its magnitude based on signal acoustics.

2. Methods

Nineteen undergraduates were recruited from the Department of Psychological and Brain Sciences at the University of Louisville. All participants reported being native English speakers with normal hearing and received course credit for their participation.

Target stimuli were a series of ten morphed natural /da/-/ga/ tokens spoken by an adult man (duration = 365 ms, 11 025 Hz sampling rate). Tokens were taken from Stephens and Holt (2011). Across the series, F_3 onset frequencies varied from 2703 Hz (/da/ endpoint) to 2338 Hz (/ga/ endpoint) before converging at/near 2614 Hz for the following /a/. The duration of the consonant transition was 63 ms.

The precursor sentence was "Correct execution of my instructions is crucial" spoken by an adult man (duration = 2200 ms, 16 000 Hz sampling rate) in the TIMIT database (Garofolo *et al.*, 1990). This sentence was selected because it had equal average energy in two frequency regions of interest: 1700–2700 Hz and 2700–3700 Hz. The frequency regions were 1000 Hz wide (after Holt, 2005, 2006). Formant transitions in stop-vowel syllables are in part defined by the formant frequencies of the following vowel, and these frequency regions reflect average F_3 frequencies for vowels spoken by men, women, and children (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995). 1700–2700 Hz includes F_3 for most vowels produced by adult men as well as /ɝ/ as spoken by adult women and children; this region is designated as "low F_3." 2700–3700 Hz captures /i/ for adult men and vowels other than /ɝ/ for adult women and children; this region is designated as "high F_3." Spectral peaks were added to the precursor sentence in either the low-F_3 or high-F_3 frequency region[2] via finite impulse response filtering with 1200 coefficients in MATLAB. Low-F_3 peaks in the precursor sentence were predicted to elicit more "da" responses, and high-F_3 peaks in the precursor sentence were predicted to elicit more "ga" responses. Following Stilp *et al.* (2015), filter gain was set to +5, +10, +15, or +20 dB. Finally, all filtered sentences and CV targets were low-pass filtered at 5 kHz, equated in root-mean-square amplitude, upsampled to a sampling rate of 44 100 Hz, and concatenated with a silent 50-ms interstimulus interval to create experimental trials. Sound examples for the precursor sentence in its unmodified form (Mm. 9) plus at all 4 filter gain values in the high F_3 region (Mm. 1–Mm. 4) and the low F_3 region (Mm. 5–Mm. 8) are given in the Supplementary Material.

Mm. 1. Precursor sentence, +5 dB peak in high F_3 region. This is a file of type "wav" (190 kB).

Mm. 2. Precursor sentence, +10 dB peak in high F_3 region. This is a file of type "wav" (190 kB).

Mm. 3. Precursor sentence, +15 dB peak in high F_3 region. This is a file of type "wav" (190 kB).

Mm. 4. Precursor sentence, +20 dB peak in high F_3 region. This is a file of type "wav" (190 kB).

Mm. 5. Precursor sentence, +5 dB peak in low F_3 region. This is a file of type "wav" (190 kB).

Mm. 6. Precursor sentence, +10 dB peak in low F_3 region. This is a file of type "wav" (190 kB).

Mm. 7. Precursor sentence, +15 dB peak in low F_3 region. This is a file of type "wav" (190 kB).

Mm. 8. Precursor sentence, +20 dB peak in low F_3 region. This is a file of type "wav" (190 kB).

Mm. 9. Precursor sentence, unfiltered. This is a file of type "wav" (190 kB).

Listeners participated individually in single-wall sound-isolating booths (Acoustic Systems, Inc., Austin, TX). Following the acquisition of informed consent, listeners completed four blocks of 160 trials in random orders. Each block tested a single level of filter gain, where each version of the precursor sentence (low F_3, high F_3) was presented with each CV target (ten-step series) eight times apiece. Blocks were tested in random orders. Listeners responded by clicking the mouse to indicate which consonant they heard on each trial. Stimuli were D/A converted by RME HDSPe AIO sound cards (Audio AG, Haimhausen, Germany) on personal computers and passed through a programmable attenuator (TDT PA4, Tucker-Davis Technologies, Alachua, FL) and headphone buffer (TDT HB6). Trial sequences were presented diotically at an average of 70 dB Sound Pressure Level via circumaural headphones (Beyerdynamic DT-150, Beyerdynamic Inc., Farmingdale, NY). The entire experiment took approximately 45 min.

3. Results

An exclusionary criterion of failing to achieve at least 80% correct on CV series endpoints was applied (Assgari and Stilp, 2015). This resulted in the removal of one listener's data, so analyses were based on responses from the remaining 18 listeners. Results were analyzed using generalized linear mixed-effect models in R (R Development Core Team, 2016) using the lme4 package (Bates *et al.*, 2014). Model architecture matched that used by Stilp *et al.* (2015), and the dependent variable was modeled as binary ("da" or "ga"). Fixed effects in the model included consonant target (coded as a continuous variable from 1 to 10 in steps of 1 then mean-centered, spanning −4.5 to +4.5), filter frequency (categorical variable with two levels: low F_3 and high F_3, with high F_3 set as the default level), filter gain (in dB; coded as a continuous variable from 5 to 20 in steps of 5 then mean-centered, spanning −7.5 to +7.5), and the interaction between filter frequency and filter gain. Random slopes were included for each main fixed effect and interaction, and a random intercept of listener was also included. The final model had the following form:

$$\begin{aligned}&\text{Response} \sim \text{Target} + \text{FilterFrequency} + \text{FilterGain} + \text{FilterFrequency} * \text{FilterGain}\\&\quad + (1 + \text{Target} + \text{FilterFrequency} + \text{FilterGain} + \text{FilterFrequency} * \text{FilterGain} \mid \text{Listener})\end{aligned}$$

Model coefficients are listed in Table 1, and predicted responses from the model are overlaid on listeners' mean responses in Fig. 1. Estimates in Table 1 are relative to the default level of Filter Frequency (high F_3) and values of 0 for mean-centered variables (Target, corresponding to the hypothetical stimulus 5.5 on the 10-step continuum; Filter Gain, corresponding to the hypothetical filter gain of 12.5 dB). The significant positive effect of Target (i.e., logistic function slope) predicts more "ga" responses in the high-F_3-filtered condition with each rightward step along the consonant target continuum (toward the /ga/ endpoint). The significant negative effect of Filter Frequency predicts a decrease in "ga" responses when the filtering condition is changed from high F_3 to low F_3, consistent with the hypothesized direction of SCEs. The significant positive effect of Filter Gain predicts an increase in "ga" responses for the high-F_3-filtered condition (i.e., more positive intercept for the high-F_3-filtered logistic function) for each 1-dB increase in filter gain. Most importantly, the interaction between filter frequency and filter gain was statistically significant. The interaction predicts that for each 1-dB increase in filter gain, listeners will respond "ga" less often when the filtering condition is changed from high F_3 to low F_3. Put another way, the model predicts that SCEs will increase as filter gain increases, which is consistent with the results depicted in Fig. 1.

As a *post hoc* analysis, following the approach of Stilp *et al.* (2015), the mixed-effects model was reanalyzed with Filter Gain coded as a categorical factor.

Table 1. Mixed-effects model results. "Target" refers to the slope of the logistic function, defined as the change in log odds of the listener responding "ga" resulting from a change of one step along the consonant continuum. "FilterFreq" lists the change in log odds of the listener responding "ga" resulting from changing the reliable spectral peak in the preceding sentence from the high F_3 region (2700–3700 Hz) to the low F_3 region (1700–2700 Hz). "FilterGain" lists the change in log odds of a "ga" response resulting from increasing peak filter gain by 1 dB. "FilterFreq × FilterGain" indicates the change in the size of the FilterFreq effect (i.e., SCE) per dB of filter gain. SE = standard error of the mean.

	Estimate	SE	Z	p
Intercept	0.36	0.24	1.49	0.14
Target	1.76	0.12	14.52	$<2\times10^{-16}$
FilterFreq	−1.86	0.25	−7.54	$<5\times10^{-14}$
FilterGain	0.07	0.02	3.99	$<7\times10^{-5}$
FilterFreq × FilterGain	−0.13	0.03	−4.45	$<9\times10^{-6}$

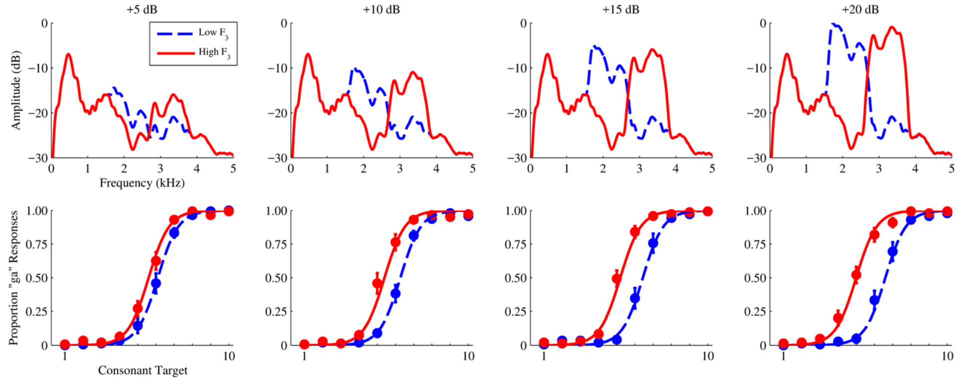

Fig. 1. (Color online) Precursor sentence spectra and mean responses. The first row depicts long-term average spectra for the filtered precursor sentences. Filter gain increases from +5 dB to +20 dB from left to right. Blue dashed lines indicate precursors with the low-F_3 region (1700–2700 Hz) amplified; red solid lines indicate precursors with the high-F_3 region (2700–3700 Hz) amplified. In the second row, symbols display mean proportions of "ga" responses (plotted along the y-axis) to each CV target (plotted along the x axis; 1 = /da/ endpoint, 10 = /ga/ endpoint) for the 18 listeners. Blue circles represent mean responses to the low-F_3-amplified precursor sentence, and red circles represent mean responses to the high-F_3-amplified precursor sentence. Error bars depict one standard error of the mean. Curves indicate the predicted logistic functions for each condition as generated by the mixed-effects model: blue dashed lines represent predicted responses in the low-F_3-amplified condition; red solid lines represent predicted responses in the high-F_3-amplified condition.

This selected one level of Filter Gain as the default level, then tested its model coefficient against 0 using a Wald z-test. All other model parameters matched those described in the previous analysis. This process was repeated for all four levels of filter gain, and predicted SCE magnitudes were derived from the model each time (Fig. 2). As in Stilp *et al.* (2015), SCE magnitude was operationalized as the distance between logistic function 50% points measured in stimulus steps along the target continuum. Given that the high-F_3 condition was used as the default level, its 50% point was calculated as −Intercept/Target. For the low-F_3 filtering condition, its 50% point was calculated as −(Intercept + Filter Frequency)/Target. All SCEs were significantly greater than 0 (all $z > 4.12$, $p < 4 \times 10^{-5}$). Critically, SCE magnitude was linearly related to filter gain. As stable spectral peaks in the precursor sentence became more prominent, SCE magnitude increased ($r = 0.99$, $p < 0.025$).[3] This replicates the linear relationship between stable spectral peaks and SCE magnitudes in vowel categorization reported by Stilp *et al.* (2015).[4]

4. Discussion

Categorization of voiced stop consonants was influenced by SCEs, as has been reported for this particular consonant contrast (Holt, 2005, 2006; Laing *et al.*, 2012)

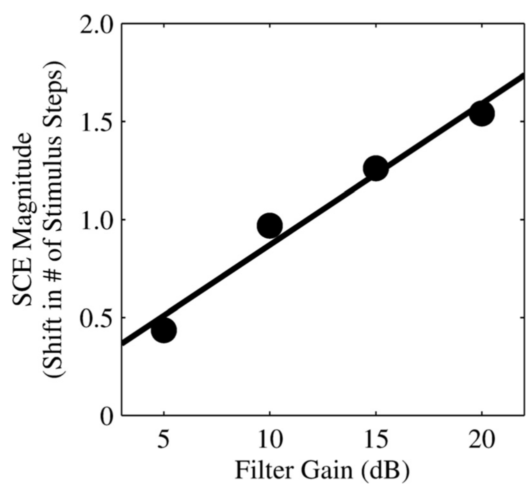

Fig. 2. Spectral contrast effects in consonant categorization are highly linear. Contrast effect magnitude is plotted on the y-axis, measured as the number of stimulus steps separating 50% points on the two logistic functions (low-F_3-amplified precursor, high-F_3-amplified precursor) in each condition. Filter gain is plotted on the x axis, indicating the prominence of the stable spectral peak added to the precursor sentence. Each circle indicates the estimated SCE magnitude for that particular filter gain as predicted by the mixed-effects model. The solid line is the linear regression fit to the data ($R^2 = 0.97$).

and others (Watkins and Makin, 1996; Sjerps and Reinisch, 2015). Importantly, the present results extend effects first reported in vowel categorization (Stilp *et al.*, 2015) to consonant categorization. In both cases, SCEs exhibited a graded influence on speech sound categorization: as the prominence of the reliable spectral peak in the preceding sentence increased, the magnitude of the resulting SCE increased in a linear fashion, producing progressively larger shifts in categorization. By generalizing this relationship across consonant and vowel classes, frequency regions (<1000 Hz in Stilp *et al.*, 2015 versus $\geq$ 1700 Hz here), and spectral cues (F_1 versus F_3), linear scaling of SCEs might be a fundamental property of speech categorization most broadly.

Here and elsewhere, SCEs have been tested by presenting different (filtered) versions of the same precursor sentence on every trial. In most cases, the precursor sentence was directly related to the listener's task ["Now you'll get (target word) to click on" (Watkins, 1991); "Please say what this vowel is" (Stilp *et al.*, 2015); etc.]. The present experiment instead utilized the sentence "Correct execution of my instructions is crucial," which had no semantic connection to the task or target items, yet SCEs were still observed. This strongly suggests that semantic content of the precursor sentence was irrelevant for producing SCEs in speech categorization. This reaffirms arguments that low-level acoustic properties play the primary role in producing SCEs and not higher-level linguistic properties of the precursor sentence (Mitterer, 2006; Laing *et al.*, 2012). This point is also supported by SCEs being produced by a different sentence on every trial (Assgari and Stilp, 2015), unintelligible time-reversed speech (Watkins, 1991), and nonspeech contexts (Holt, 2005, 2006; Laing *et al.*, 2012).

The present results may offer important insights for understanding and ameliorating speech perception by listeners with hearing impairment. Historically, research on SCEs examined normal-hearing listeners exclusively, but recent efforts suggest that this phenomenon extends to hearing-impaired listeners as well. Listeners with sensorineural hearing loss not only experienced SCEs in their vowel categorization, but significantly larger shifts in categorization than those observed for normal-hearing listeners (Stilp and Alexander, 2016). The same may be true for cochlear implant users, as acoustic simulations of cochlear implant processing also produced significantly larger SCEs (Stilp, 2017). This is problematic because when SCEs are too large, phoneme categories are shifted too far apart from each other, making previously unambiguous sounds more perceptually ambiguous and thus misidentified (see Stilp and Alexander, 2016 and Stilp, 2017 for discussions). SCEs are pervasive in normal-hearing listeners' perception of consonants and vowels, raising the question of whether the same is true for listeners with impaired hearing. If this is indeed the case, how spectral context is processed by hearing aids and cochlear implants merits close consideration in order to make the influence of spectral context more comparable to that for normal-hearing listeners.

In conclusion, a close coupling was demonstrated between signal acoustics and speech categorization. Stable spectral properties of the acoustic context were linearly related to the size of the resulting SCE, a pattern which was first reported for vowel categorization (Stilp *et al.*, 2015) and replicated here in consonant categorization. In both cases, this relationship persisted even for very subtle manipulations of the precursor sentence (+5 dB filter gain). In generalizing across consonants and vowels, frequency regions, and spectral cues, acute sensitivity to spectral context may well be fundamental to speech perception most broadly.

Acknowledgments

The authors thank Ashley Batliner, Alexandrea Beason, Carly Newman, and Madison Rhyne for their assistance with data collection.

References and links

[1]There are two classes of SCEs in the speech perception literature. The first class defines the acoustic context as the sound immediately before the target sound (e.g., Lotto and Kluender, 1998). The second class, to which the present report belongs, defines the acoustic context as a longer duration of sounds before the target sound, such as a sentence. Predictions and effect directions are highly consistent across the two classes of SCEs. The primary difference is whether the key spectral property in the preceding context (which is responsible for producing the SCE) is restricted to its offset or is instead relatively stable or recurring throughout the context.

[2]This is a departure from earlier studies by Holt (2005, 2006) and Laing *et al.* (2012), where frequency regions of interest (1300–2300 and 2300–3300 Hz) were based on pure tone stimuli used by Lotto and Kluender (1998) and were simply designated as "low" and "high".

[3]The reported correlation used the model prediction for SCE magnitude at each level of filter gain. When SCEs were calculated for each listener at each level of filter gain using individual logistic regressions, mean SCEs were correlated with filter gain equally strongly ($r = 0.99$, $p < 0.025$).

[4]Stilp *et al.* (2015) calculated correlations between different filter characteristics and SCE magnitudes. While peak filter gain was a significant predictor of SCE magnitude, total filter power (root-mean-squared amplitude of filter gain plotted as a function of frequency) was a superior predictor. This difference in predictive power was because correlations were calculated across various filter types (100-Hz bandwidth spectral peak, 300-Hz bandwidth spectral peak, 2500-Hz bandwidth spectral envelope difference), dissociating peak filter gain from its total power. Here, a single filter type was used (1000-Hz bandwidth spectral peak), so correlating SCE magnitudes with peak filter gain ($r = 0.9860$, $p = 0.0140$) versus total filter power ($r = 0.9857$, $p = 0.0143$) produced nearly identical results.

Ainsworth, W. (**1975**). "Intrinsic and extrinsic factors in vowel judgments," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. Tatham (Academic Press, London), pp. 10–113.
Assgari, A. A., and Stilp, C. E. (**2015**). "Talker information influences spectral contrast effects in speech categorization," J. Acoust. Soc. Am. **138**(5), 3023–3032.
Bates, D. M., Maechler, M., Bolker, B., and Walker, S. (**2014**). "lme4: Linear mixed-effects models using Eigen and S4. R package (version 1.1-7)," http://cran.r-project.org/package=lme4 (Last viewed 2/3/2017).
Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (**1962**). "The identification and discrimination of synthetic vowels," Language Speech **5**(4), 171–189.
Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (**1990**). "DARPA TIMIT acoustic-phonetic continuous speech corpus CDROM," NIST Order No. PB91-505065, National Institute of Standards and Technology, Gaithersburg, MD.
Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5), 3099–3111.
Holt, L. L. (**2005**). "Temporally nonadjacent nonlinguistic sounds affect speech categorization," Psychol. Sci. **16**(4), 305–312.
Holt, L. L. (**2006**). "The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization," J. Acoust. Soc. Am. **120**(5), 2801–2817.
Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**(1), 98–104.
Laing, E. J., Liu, R., Lotto, A. J., and Holt, L. L. (**2012**). "Tuned with a tune: Talker normalization via general auditory processes," Front. Psychol. **3**, 1–9.
Lotto, A. J., and Kluender, K. R. (**1998**). "General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification," Percept. Psychophys. **60**(4), 602–619.
Mitterer, H. (**2006**). "Is vowel normalization independent of lexical processing?" Phonetica **63**(4), 209–229.
Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**(5), 2088–2113.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**(2), 175–184.
Pisoni, D. B. (**1973**). "Auditory and phonetic memory codes in the discrimination of consonants and vowels," Percept. Psychophys. **13**(2), 253–260.
R Development Core Team. (**2016**). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org/ (Last viewed 2/3/2017).
Sjerps, M. J., Mitterer, H., and McQueen, J. M. (**2011**). "Constraints on the processes responsible for the extrinsic normalization of vowels," Percept. Psychophys. **73**(4), 1195–1215.
Sjerps, M. J., and Reinisch, E. (**2015**). "Divide and conquer: How perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception," J. Exp. Psychol.: Human Perc. Perform. **41**(3), 710–722.
Stephens, J. D. W., and Holt, L. L. (**2011**). "A standard set of American-English voiced stop-consonant stimuli from morphed natural speech," Speech Commun. **53**(6), 877–888.
Stilp, C. E. (**2017**). "Reliable spectral properties elicit contrast effects in perception of noise-vocoded speech," Assoc Res Otolaryngol Abs (in press).
Stilp, C. E., and Alexander, J. M. (**2016**). "Spectral contrast effects in vowel categorization by listeners with sensorineural hearing loss," Proc. Meet. Acoust. **26**, 060003.
Stilp, C. E., Anderson, P. W., and Winn, M. B. (**2015**). "Predicting contrast effects following reliable spectral properties in speech perception," J. Acoust. Soc. Am. **137**(6), 3466–3476.
Watkins, A. J. (**1991**). "Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion," J. Acoust. Soc. Am. **90**(6), 2942–2955.
Watkins, A. J., and Makin, S. J. (**1996**). "Some effects of filtered contexts on the perception of vowels and fricatives," J. Acoust. Soc. Am. **99**(1), 588–594.