

Research Paper

Speech perception adjusts to stable spectrotemporal properties of the listening environment

Christian E. Stilp^{a,*}, Paul W. Anderson^b, Ashley A. Assgari^a, Gregory M. Ellis^a, Pavel Zahorik^a^a Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY, 40292, USA^b Department of Psychology, Murray State University, Murray, KY, 42071, USA

ARTICLE INFO

Article history:

Received 15 March 2016

Received in revised form

10 August 2016

Accepted 15 August 2016

Available online 3 September 2016

Keywords:

Speech perception

Spectral calibration

Perceptual constancy

Reverberation

ABSTRACT

When perceiving speech, listeners compensate for reverberation and stable spectral peaks in the speech signal. Despite natural listening conditions usually adding both reverberation and spectral coloration, these processes have only been studied separately. Reverberation smears spectral peaks across time, which is predicted to increase listeners' compensation for these peaks. This prediction was tested using sentences presented with or without a simulated reverberant sound field. All sentences had a stable spectral peak (added by amplifying frequencies matching the second formant frequency [F_2] in the target vowel) before a test vowel varying from /i/ to /u/ in F_2 and spectral envelope (tilt). In Experiment 1, listeners demonstrated increased compensation (larger decrease in F_2 weights and larger increase in spectral tilt weights for identifying the target vowel) in reverberant speech than in nonreverberant speech. In Experiment 2, increased compensation was shown not to be due to reverberation tails. In Experiment 3, adding a pure tone to nonreverberant speech at the target vowel's F_2 frequency increased compensation, revealing that these effects are not specific to reverberation. Results suggest that perceptual adjustment to stable spectral peaks in the listening environment is not affected by their source or cause.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Much in the sensory environment is predictable from time to time and place to place. Sensory systems have adapted and evolved to be sensitive to this predictability (Attneave, 1954; Barlow, 1961). Sensitivity to stable aspects of the environment promotes a host of perceptual phenomena: grouping, scene analysis, and source localization, among others. The auditory system responds to predictability in the sensory environment through several related mechanisms: adaptation, constancy, normalization, compensation, and calibration. While some of these mechanisms may differ only in name or in scale, they all serve audition by providing adjustment to stable aspects of the listening environment. Compensating for regularities in the sensory environment can affect processing of simple acoustic properties, such as adapting to a particular

frequency or entraining to a regular rhythm. Environmental regularities can also affect higher-level auditory processing, such as sound source identification, speech understanding, and object recognition. The present focus is on stable spectrotemporal properties of the listening environment that influence speech perception.

In everyday perception, sounds are filtered by the listening environment. As sounds propagate from source to perceiver, different frequencies are amplified or attenuated depending on the composition of the listening environment. This filtering can make certain frequencies particularly prominent and relatively stable across time, producing *stable spectral properties*. Several reports have shown that perception deemphasizes these stable properties and increases reliance on changing (less predictable, and thus more informative) signal properties (Kiefte and Kluender, 2008; Alexander and Kluender, 2010; Stilp and Anderson, 2014). For example, when earlier sounds feature a stable spectral peak that matches the second formant frequency (F_2) of the following target vowel (perceptually varying from /i/ to /u/, for which F_2 is a key distinguishing feature), listeners decreased their reliance on F_2 and

* Corresponding author.

E-mail addresses: christian.stilp@louisville.edu (C.E. Stilp), panderson9@murraystate.edu (P.W. Anderson), ashley.assgari@louisville.edu (A.A. Assgari), g.ellis@louisville.edu (G.M. Ellis), pavel.zahorik@louisville.edu (P. Zahorik).

increased their reliance on changing, more informative cues for vowel identification such as spectral tilt. In the literature, this process has been called *auditory perceptual calibration* (Kiefte and Kluender, 2008; Alexander and Kluender, 2010; Stilp and Anderson, 2014). Here we adopt the more descriptive term *spectral calibration* to distinguish it from calibration to other stable acoustic properties or contingencies between properties. Spectral calibration is a key mechanism for factoring out predictable acoustic aspects of the listening environment, and is analogous to color constancy in vision (see Alexander and Kluender, 2010; Stilp et al., 2010 for discussions).

Natural listening environments also produce acoustic reflections that interact with the source signal. Reverberant acoustic energy can degrade the intelligibility and quality of speech, especially when reverberation times are long (Knudsen, 1929; Nábělek and Robinson, 1982; Nábělek and Donahue, 1984; Nábělek and Letowski, 1985). However, given sufficient exposure, listeners compensate for stable patterns of reverberation (Watkins, 2005a, 2005b; Watkins et al., 2011; Watkins and Makin, 2007; Brandewie and Zahorik, 2010, 2013; Srinivisan and Zahorik, 2013, 2014). This compensation has been shown to improve speech intelligibility considerably (Brandewie and Zahorik, 2010, 2013; Srinivisan and Zahorik, 2013, 2014). From this perspective, compensating for reverberation is another instance of perceptual constancy in speech perception (Assmann and Summerfield, 2004; Watkins and Makin, 2007). While not traditionally viewed as such, a given pattern of reverberation can serve as a *stable spectrotemporal property* of the listening environment, producing characteristic spectrotemporal alterations to the source signal.

Listening environments alter the frequency compositions of sounds while also producing acoustic reflections. Listeners often factor out these stable properties of the acoustic environment, whether they are primarily spectral (as in spectral calibration) or spectrotemporal (as in compensation for reverberation). In terms of perceptual adjustment to stable properties of the listening

environment, compensations for stable spectral properties and reverberation are highly related. Intriguingly, these processes offer separate notions of what makes a particular spectral property “stable”, whether it is prominence in the spectral domain (e.g., relative amplitude of a spectral peak, stability of overall spectral shape) or the spectrotemporal domain (e.g., patterns of temporal elongation due to acoustic reflections).

Despite their broad similarities and likely co-occurrence in speech perception, spectral calibration and compensation for reverberation have been studied separately using different tasks. Compensation for reverberation has been studied by measuring speech intelligibility or word recognition, while spectral calibration studies have examined the perceptual weighting of spectral cues for vowel categorization. While speech intelligibility and cue weighting are not unrelated (Winn and Litovsky, 2015), they are sufficiently distinct to obscure the relative contributions of spectral calibration and compensation for reverberation to speech perception.

In studies of spectral calibration, stable spectral peaks were added to a preceding acoustic context through filtering. This filtered context featured a stable spectral peak that matched the F_2 center frequency in the subsequent target vowel, and identification of this vowel was altered by this spectral peak in earlier sounds. Previous investigations used filters with narrow bandwidths (100 Hz; Fig. 1a) in order to minimally affect speech quality and intelligibility. However, reverberation can significantly impair speech quality in ways that produce clear predictions for how spectral cue use would differ in reverberant and nonreverberant listening conditions. Reverberation acts as a low-pass filter in the amplitude modulation domain (Houtgast and Steeneken, 1973). Stable spectral peaks wax and wane along with speech energy in the passband region (generally between 3 and 8 Hz; Houtgast and Steeneken, 1985; Elliott and Theunissen, 2009) and would be smeared across time in reverberant listening conditions (Fig. 1b). Hearing stable spectral properties more often in a fixed amount of

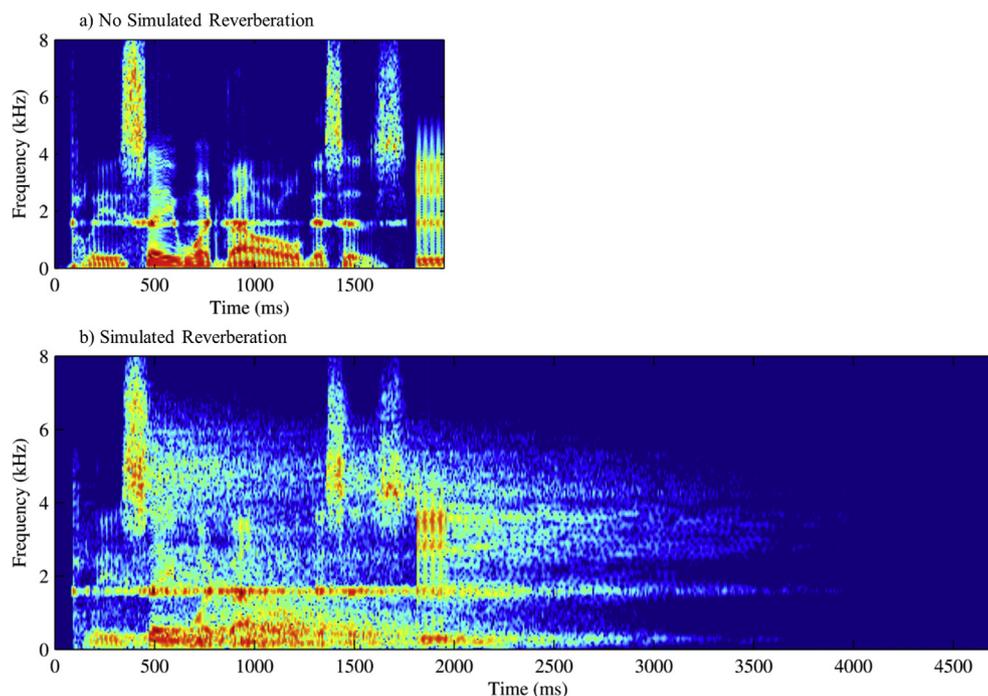


Fig. 1. Spectrograms for sample trials from Experiment 1. Both precursor sentences (“Please say what vowel this is”) featured a stable spectral peak at 1600 Hz, followed by a target vowel with $F_2 = 1600$ Hz and spectral tilt = -3 dB/octave. (a) Sentence without simulated reverberation, Experiment 1a. (b) Sentence in a simulated reverberant sound field (broadband $T_{60} = 3476$ ms), Experiment 1b. Spectrograms are time-aligned to illustrate longer trial durations for reverberant stimuli.

time has been shown to increase the degree of spectral calibration (*i.e.*, there is a larger decrease in perceptual weight for the stable spectral cue and a larger increase in weight for the changing spectral cue; Alexander and Kluender, 2010). Thus, spectral calibration is predicted to increase in reverberant listening conditions relative to nonreverberant listening conditions.

In the present experiments, listeners first identified isolated target vowels as “ee” (/i/) or “oo” (/u/). These vowels varied along two spectral dimensions, one narrow (center frequency of F_2) and one broad (spectral tilt, or overall spectral envelope shape). These dimensions form a trading relation for perception of /i/ and /u/, which is ideal for investigating perceptual compensation to a listening context where listeners emphasize one cue and deemphasize the other cue. Listeners then identified these same vowels when presented after a precursor sentence. Precursor sentences had increased energy near the center frequency of F_2 in the target vowel, produced either via selective filtering (Experiments 1a, 1b, 2) or via concurrent presentation of a tone (Experiment 3). Sentences were then processed using simulated room reverberation to study its effect on the degree of spectral calibration (Experiments 1b, 2). Cue weights (standardized logistic regression coefficients) were used to estimate listeners’ reliance on F_2 and spectral tilt for identifying vowels in isolation versus following the precursor sentence.

Consistent with previous investigations of spectral calibration (Kiefte and Kluender, 2008; Alexander and Kluender, 2010; Stilp and Anderson, 2014), when the preceding acoustic context shares a spectral peak with F_2 of the target vowel, listeners are predicted to weight F_2 less (relative to F_2 weight for identifying isolated vowels) and to weight tilt more (relative to tilt weight for identifying isolated vowels). Given that reverberation smears stable spectral energy in the precursor sentence across time, listeners are predicted to display greater spectral calibration (larger weight changes) in reverberant listening conditions than in nonreverberant listening conditions. The present experiments tested this prediction and explored its underlying sources.

2. Experiment 1

2.1. Methods

2.1.1. Listeners

Forty undergraduate students were recruited from the Department of Psychological and Brain Sciences at the University of Louisville. Twenty listeners participated in Experiment 1a, and 20 different listeners participated in Experiment 1b. In all experiments, listeners were native English speakers with self-reported normal hearing, and received course credit for their participation. All procedures involving human listeners were approved by the University of Louisville Institutional Review Board.

2.1.2. Stimuli

2.1.2.1. Vowels. Target vowels were the same stimuli as used by Alexander and Kluender (2010) and Stilp and Anderson (2014). Vowels were synthesized using the parallel branch of the Klatt and Klatt (1990) synthesizer with a fundamental frequency of 100 Hz and 146-ms duration (5-ms linear onset/offset ramps). A series of five vowels varying from /u/ (as in “boot”) to /i/ (as in “beet”) was created by varying the second formant frequency (F_2) from 1000 Hz to 2200 Hz in 300-Hz steps. F_2 bandwidth (160 Hz) and the center frequencies and bandwidths of other formants were held constant (F_1 : 300 Hz center frequency, 160 Hz bandwidth; F_3 : 2700 Hz center frequency, 260 Hz bandwidth; F_4 : 3600 Hz center frequency, 360 Hz bandwidth). Presenting vowels with steady formant frequencies avoided confusions where reverberant diphthongal

vowels are misidentified as their initial vowel (Nábělek and Letowski, 1985; Nábělek and Dagenais, 1986). Formant amplitudes were manipulated to ensure that each vowel had a reasonably constant spectral tilt of -3 dB/octave, as measured by linear regression slope across the log-frequency spectrum.

Spectral tilt was then manipulated using 90-tap finite impulse response filters in MATLAB (Mathworks Inc., Natick, MA). Between 212 and 4800 Hz, filter gain changed linearly in dB as a function of log frequency. The tilt of the filter response varied from -9 to $+3$ dB/octave in 3 dB/octave steps. Each of these five tilt filters was applied to each member of the five-step F_2 series of vowels described above, creating a fully crossed 5-by-5 vowel matrix where F_2 varied from 1000 to 2200 Hz in 300-Hz steps and final spectral tilt varied from -12 to 0 dB/octave in 3 dB/octave steps. Vowel targets were then low-pass filtered using a finite impulse response filter with an upper cutoff at 4800 Hz and stopband of -90 dB at 6400 Hz.

2.1.2.2. Precursor sentence. The precursor sentence was “Please say what vowel this is” produced by AT&T Natural.

Voices™ Text-to-Speech Synthesizer (Beutnagel et al., 1997). The male talker (“Mike”) had an American English accent, and was the same token as used by Stilp and Anderson (2014). The precursor duration was 1759 ms.

The precursor was processed by a filter centered at one of the F_2 frequencies used in the vowel matrix (1000, 1300, 1600, 1900, and 2200 Hz). The filter gain was set to $+20$ dB over the range ± 50 Hz around the center frequency and 0 dB elsewhere. Filters were created using the fir2 function in MATLAB with 1200 coefficients. Target vowels and filtered precursors had equal RMS amplitude. Target vowels were then appended to filtered precursors sharing the same spectral peak (F_2 in the vowel) separated by a 50-ms silent inter-stimulus interval (ISI).

2.1.2.3. Reverberant sound field simulation. For Experiments 1b and 2, virtual acoustic modeling techniques, identical to those used by Zahorik (2009), were used to simulate the acoustics of a reverberant room over headphones. As in Brandewie and Zahorik (2010), an equalization filter was applied to correct for the loudspeaker response used during head-related transfer function (HRTF) measurement procedures. Non-individualized HRTF measurements from a single listener were used to spatially render the direct-path and early reflections. The modeling techniques simulated early reflections using an image model (Allen and Berkley, 1979) while a statistical model simulated late reverberation. This modeling technique has been found to generate binaural room impulse responses (BRIRs) that are reasonable physical and perceptual approximations of those measured in a real room (Zahorik, 2009).

For this experiment, room parameters in the model were similar to those of Room 3 tested by Brandewie and Zahorik (2010). The broadband reverberation time (T_{60}) was 3476 ms (ISO-3382, 1997). Reverberation was applied to isolated vowels and to entire sentence-plus-vowel trial sequences. For each stimulus, the left channel was selected and duplicated across ears for diotic presentation. While this might not reflect dichotic perception of reverberation in natural listening conditions, it was done to facilitate comparisons to results for diotically presented nonreverberant stimuli in Experiments 1a and 3. In reverberation at the listener’s location, the vowel duration and trial duration were increased by the reverberation time of the room (see Fig. 1b).

2.1.3. Procedure

Nonreverberant and reverberant stimuli were used in Experiment 1a and 1b, respectively, but the procedures across experiments were identical. Stimuli were resampled at 44,100 Hz

sampling rate and presented diotically at an average sound pressure level (SPL) of approximately 70 dB via circumaural headphones (Beyer-Dynamic DT-150, Beyerdynamic Inc. USA, Farmingdale, NY). Listeners participated individually in single-wall sound-isolating booths (Acoustic Systems, Inc., Austin, TX). Listeners responded by clicking the mouse to indicate whether the target vowel sounded more like “ee” or “oo.” No feedback was provided. Following acquisition of informed consent, listeners first completed a block of 200 trials (25 vowels repeated eight times) where vowels were presented in isolation. This first block is referred to as the baseline, as it permitted calculation of spectral cue weights without any influence of the preceding sentence context. Listeners then completed 200 trials where the precursor sentence was filtered to have a spectral peak matching the F_2 of the following target vowel (8 repetitions of 25 vowels, each paired with their F_2 -matched precursor sentence). This second block is referred to as the test block, as responses were influenced by the filtered precursor sentence. The entire session lasted approximately 30 min.

2.2. Results

Multiple logistic regression was used on individual listener data to predict the identified vowel category (either “ee” or “oo”) from the F_2 and spectral tilt variables. Following previous studies of spectral calibration (Kiefte and Kluender, 2008; Alexander and Kluender, 2010; Stilp and Anderson, 2014), the standardized logistic regression coefficients were taken as estimates of perceptual weights for F_2 and spectral tilt. Regressions were calculated for vowels presented in isolation to produce baseline weights and for vowels following the filtered precursor sentence to produce test weights. Spectral calibration was defined as the change in perceptual weights across the two sessions (*i.e.*, perceptual adjustment to the precursor).

Following Alexander and Kluender (2010) and Stilp and Anderson (2014), Wilcox's (2005) Minimum Generalized Variance method was used to remove all data for listeners whose weights for either session were identified as outliers. This resulted in the removal of three complete datasets from Experiment 1a and four complete datasets from Experiment 1b, resulting in responses from 17 and 16 listeners being analyzed, respectively. Group psychometric functions for remaining listeners are presented in Fig. 2. Changes in spectral cue weights are evident when comparing functions for the test session (Fig. 2b and d) to functions for the baseline session (2a and 2c). The decrease in F_2 weight is evident in shallower psychometric function slopes, whereas the increase in spectral tilt weight is apparent in more positive intercepts (*i.e.*, leftward shifts of the functions).

Mean weights and weight changes for Experiments 1a and 1b are presented in Table 1. Statistical analyses were conducted using paired-sample, independent-samples, and one-sample *t*-tests where appropriate. Tests were two-tailed unless noted as one-tailed when testing a directional prediction. Mean baseline weights in Experiment 1a were consistent with other investigations of spectral calibration, showing higher weights for F_2 than tilt for vowel identification (paired-samples *t*-test: $t_{16} = 2.84$, $p < 0.025$). Baseline weights in Experiment 1b also followed this pattern but with even greater asymmetry ($t_{15} = 11.84$, $p < 0.001$). Independent-samples *t*-tests indicated that listeners weighted F_2 significantly higher for reverberant vowels than for nonreverberant vowels ($t_{31} = 2.52$, $p < 0.025$). The weights for tilt did not differ significantly ($t_{31} = 1.65$, $p = 0.11$). Test weights, on the other hand, were similar across spectral cues and across listener groups.

F_2 weights were predicted to decrease across sessions while tilt weights were predicted to increase. One-tailed *t*-tests against zero

weight change confirmed this pattern for both Experiment 1a (F_2 weight decrease: $t_{16} = 3.77$, $p < 0.001$; tilt weight increase: $t_{16} = 2.11$, $p < 0.05$) and Experiment 1b (F_2 weight decrease: $t_{15} = 10.27$, $p < 0.001$; tilt weight increase: $t_{15} = 6.80$, $p < 0.001$). Critically, one-tailed independent-samples *t*-tests confirmed the prediction that the decrease in F_2 weights ($t_{31} = 2.74$, $p < 0.001$) and the increase in tilt weights ($t_{31} = 2.45$, $p < 0.025$) would be larger for reverberant than for nonreverberant stimuli.

2.3. Discussion

For reverberant and nonreverberant stimuli, listeners compensated for stable spectral properties in the listening context. In both cases, listeners decreased perceptual weights for the stable spectral property (F_2) and increased weights for the changing spectral property (spectral tilt). Importantly, the degree of this adjustment was significantly larger in reverberant listening conditions. This suggests that listeners alter spectral cue usage when reverberation and spectral coloration are present.

The long reverberation time introduced a constellation of acoustic differences between the reverberant and nonreverberant stimuli, inviting closer examination of what specifically produced the increased compensation. Hearing the stable spectral peaks more often due to reverberation smearing these spectral peaks across time is one possible explanation (see Fig. 1). A second, non-exclusive possibility is that reverberant stimuli simply have more energy in that frequency region than nonreverberant stimuli. This would make reverberant stimuli more effective at adapting neural responses to these frequencies in the precursor sentence and (especially) F_2 in the target vowel.

These and potentially other explanations raise the question as to whether the results were due to reverberation *per se* or to signal properties that co-occur with but are not exclusive to reverberation. Reverberation could be manipulated in a number of ways to address this question (reverberation time, source distance, direct-to-reverberant energy ratio, *etc.*), but these manipulations would change the temporal characteristics of the stable spectral peaks in Experiment 1b, complicating comparisons of results across experiments. Instead, one might decrease perceived level of reverberation by manipulating the reverberation tails. Reverberation tails promote perceptual compensation for reverberation (Watkins, 2005a, 2005b), and truncating reverberant tails at the end of test items reduces perceptual compensation for reverberation (Watkins and Raimond, 2013; Beeston et al., 2014). Removing reverberant tails from the test vowels altogether should decrease perceived reverberation while preserving the temporal characteristics of the spectral peaks in isolated vowels (relative to nonreverberant vowels in Experiment 1a) and in the precursor sentence (relative to reverberant stimuli in Experiment 1b).

In Experiment 2, the reverberant tails were removed from the test vowels so that isolated vowels and trial sequences matched the durations used in Experiment 1a. If the results from Experiment 1 were principally due to perceived reverberation, then the degree of spectral calibration in Experiment 2 should be less than that measured for the highly-reverberant stimuli in Experiment 1b. Spectral calibration might decrease if the results were instead due to temporally extended spectral peaks, as deleting reverberant tails from the test vowels decreases the overall duration of these peaks.

3. Experiment 2

3.1. Methods

3.1.1. Listeners

Twenty-three undergraduate students were recruited from the

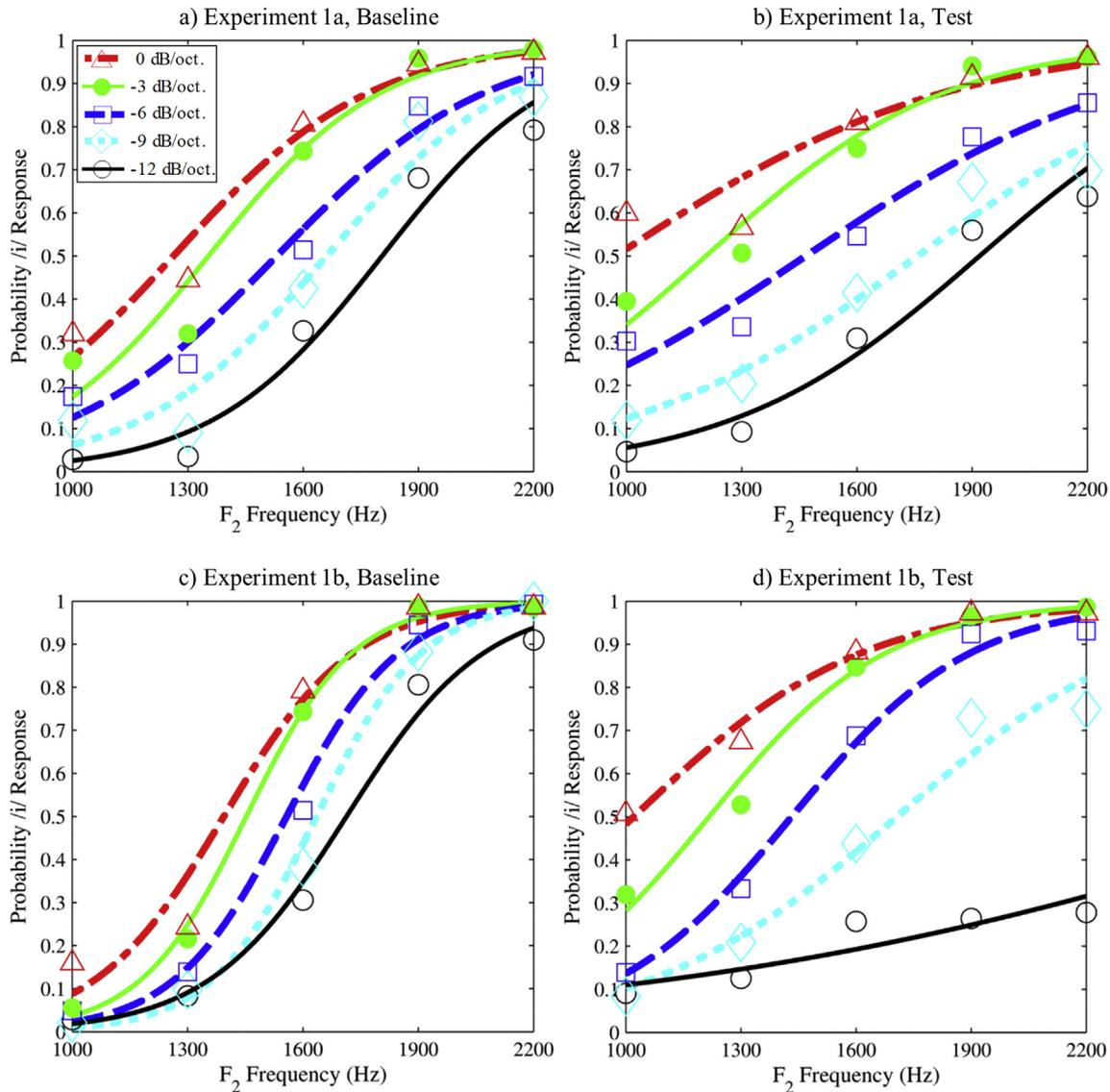


Fig. 2. Group psychometric functions from Experiment 1. The probability of responding “ee” (/i/) to the target vowel is presented as a function of the vowel F_2 frequency. The mean responses to vowels with spectral tilts of -12 , -9 , -6 , -3 , and 0 dB/octave are indicated by circles, diamonds, squares, dots, and triangles, respectively. Logistic regressions were fitted to the group data at each level of spectral tilt for illustration purposes. (a) Mean responses to isolated vowels without simulated reverberation in Experiment 1a. (b) Mean responses to vowels following the filtered precursor sentence without simulated reverberation in Experiment 1a. (c) Mean responses to isolated vowels processed with simulated reverberation in Experiment 1b. (d) Mean responses to vowels following the filtered precursor sentence in simulated reverberation in Experiment 1b.

Department of Psychological and Brain Sciences at the University of Louisville.

3.1.2. Stimuli

The same stimuli from Experiment 1b were used in Experiment 2 with only one change. Reverberation tails were removed from the target vowels, truncating the stimulus duration to match that of nonreverberant stimuli in Experiment 1a. Isolated vowels were

truncated to 146 ms duration, and trial sequences (filtered precursor sentence followed by target vowel) were truncated to 1955 ms duration (Fig. 3a).

3.1.3. Procedure

Experiment 2 followed the same procedures as Experiment 1. The entire session lasted approximately 20 min.

Table 1

Results of Experiments 1a and 1b. Mean weights (standardized logistic regression coefficients; Baseline, Test columns) and mean weight changes (Calibration columns) are presented with one standard error of the mean indicated in parentheses.

Experiment	Baseline		Test		Calibration	
	F_2	Tilt	F_2	Tilt	ΔF_2	ΔTilt
1a ($n = 17$)	1.61 (0.17)	0.68 (0.17)	1.15 (0.18)	1.03 (0.22)	-0.46 (0.12)	$+0.35$ (0.16)
1b ($n = 16$)	2.10 (0.10)	0.37 (0.08)	1.23 (0.11)	1.23 (0.14)	-0.87 (0.08)	$+0.86$ (0.13)

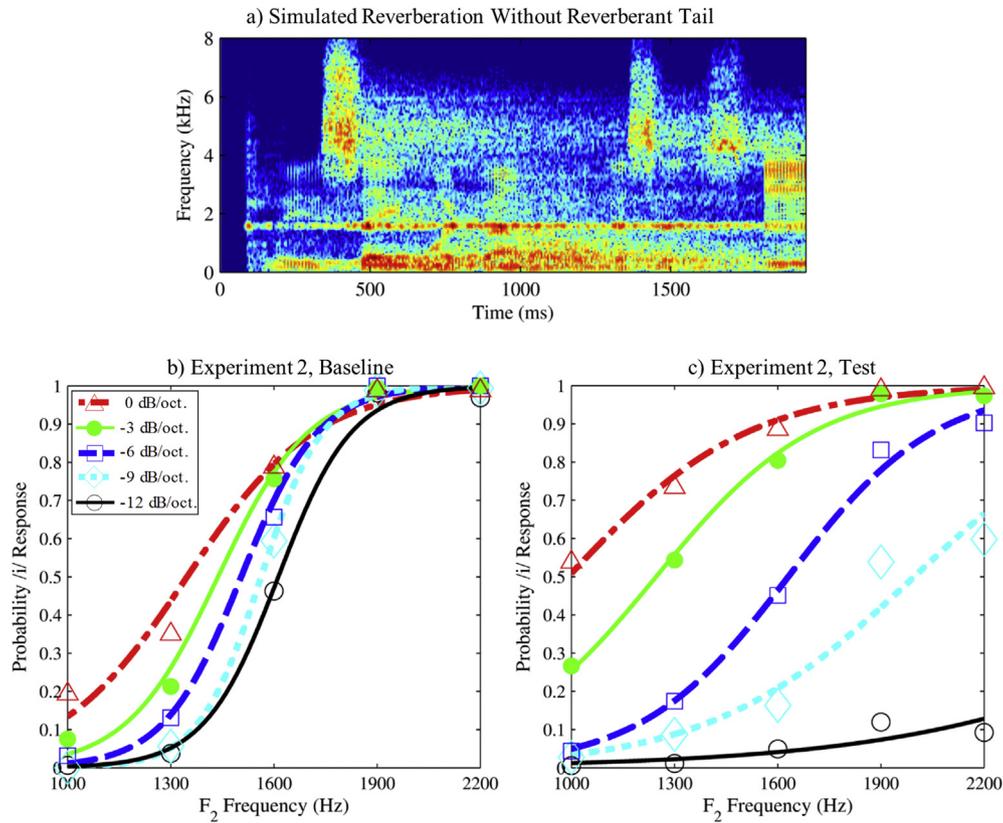


Fig. 3. (a) Spectrogram for a sample trial from Experiment 2. The stimuli matched those tested in Experiment 1b (and illustrated in Fig. 1b) but with the reverberant tail deleted so that the trial duration matched that in Experiment 1a. (b–c) Group psychometric functions from Experiment 2, following the layout of Fig. 2. Logistic regressions were fitted to the group data at each level of spectral tilt for illustration purposes. (b) Mean responses to isolated reverberant vowels without reverberation tails in Experiment 2. (c) Mean responses to reverberant vowels without reverberant tails following the precursor sentence in Experiment 2.

3.2. Results

Three listeners' results were identified as outliers according to Wilcox's (2005) Minimum Generalized Variance method, so their data sets were excluded from further analysis. Psychometric functions for group data across the remaining 20 listeners are presented in Fig. 3. Mean weights and weight changes are presented in Table 2. Listeners again weighted F₂ more heavily for vowels in isolation (paired-samples *t*-test: $t_{19} = 13.57, p < 0.001$). Baseline weights did not significantly differ from those in Experiment 1b (independent-samples *t*-tests all $t < 1.10, p > 0.28$). At test, tilt weights were similar to F₂ weights (paired-samples *t*-test: $t_{19} = 1.72, p = 0.10$) and significantly higher than tilt weights for fully reverberant stimuli in Experiment 1b (independent-samples *t*-test: $t_{34} = 2.35, p < 0.025$). F₂ test weights were similar across Experiments 1b and 2.

In terms of spectral calibration, one-tailed *t*-tests against zero again revealed weight changes in the predicted directions (F₂ weight decrease: $t_{19} = 5.84, p < 0.001$; tilt weight increase: $t_{19} = 11.05, p < 0.001$). In testing the prediction that spectral calibration would be smaller for reverberant stimuli without tails than

for stimuli with tails, one-tailed independent-samples *t*-tests revealed that decreases in F₂ weights did not differ ($t_{34} = 1.08, p = 0.14$) but increases in tilt weights were unexpectedly larger for stimuli without reverberation tails ($t_{34} = 3.12, p = 0.99$ for the one-tailed test in the predicted direction; $p < 0.01$ for a two-tailed test). Finally, one-tailed independent-samples *t*-tests confirmed the prediction of larger tilt weight increases for reverberant stimuli in Experiment 2 than nonreverberant stimuli in Experiment 1a ($t_{35} = 5.26, p < 0.001$), but the decrease in F₂ weights was not significantly larger in Experiment 2 than in Experiment 1a ($t_{35} = 1.42, p = 0.08$).

3.3. Discussion

Spectral calibration was consistent across vowel stimuli with and without reverberant tails, suggesting that the reverberant tails were not responsible for the increased effects observed in Experiment 1b. The results are contrary to those of Watkins and colleagues, who found that truncated reverberant tails reduced perceptual compensation for reverberation (Watkins, 2005a, 2005b; Watkins and Raimond, 2013; Beeston et al., 2014). Task

Table 2

Results of Experiment 2. Mean weights (standardized logistic regression coefficients; Baseline, Test columns) and mean weight changes (Calibration columns) are presented with one standard error of the mean indicated in parentheses.

Experiment	Baseline		Test		Calibration	
	F ₂	Tilt	F ₂	Tilt	ΔF ₂	ΔTilt
2 (<i>n</i> = 20)	1.96 (0.08)	0.27 (0.05)	1.26 (0.13)	1.70 (0.14)	-0.70 (0.12)	+1.43 (0.13)

differences might be responsible for this difference in results. The present experiments measured spectral cue weights for identifying vowels, whereas Watkins and colleagues measured shifts in phoneme category boundaries along a continuum that changed from “sir” to “stir” based on the duration of a temporal gap. Additionally, Watkins (2005b) proposed that reverberant tails at the end of the signal as well as those filling in regions of spectral transitions can aid compensation for reverberation, but only the former type were manipulated here.

Surprisingly, tilt weights were significantly higher at test in Experiment 2 than in Experiment 1b. Deleting the reverberant tail might have helped listeners perceive and use spectral tilt to label the target vowel, as they no longer had to wait for the reverberant tail to decay before responding. However, this explanation is challenged by similar tilt weights for isolated vowels in Experiment 1b and Experiment 2. The reason for the significant increase in tilt weights is not clear, but it should be remembered that different listeners completed Experiments 1b and 2.

While the results of Experiment 2 suggest that increased spectral calibration is not due to the presence of reverberation tails, it does not distinguish whether the effects are due to the reverberation *per se* or simply acoustic byproducts of reverberation. All results thus far are consistent with a total energy model, where greater energy in F_2 regions of reverberant stimuli elicits larger calibration than for nonreverberant stimuli. This increased energy is due to reverberation acting as a low-pass filter in the modulation domain (Houtgast and Steeneken, 1973), smearing stable spectral energy across time and increasing its effective duration (Figs. 1b and 3a). Signal manipulations that mimic these effects without introducing reverberation would better indicate whether spectral

calibration increases specifically when reverberation is present.

In Experiment 3, stable spectral peaks were introduced in nonreverberant stimuli by adding a pure tone that matched the F_2 center frequency of the target vowel. This produced a temporally extended stable spectral peak in the precursor sentence. If the increased spectral calibration in Experiments 1b and 2 were due to reverberation specifically, then spectral calibration in Experiment 3 should not increase and would likely approximate that for the nonreverberant stimuli tested in Experiment 1a. If the results from Experiments 1b and 2 were instead due to increased energy at the spectral peaks in the precursor sentence, then increased calibration should be replicated in Experiment 3.

4. Experiment 3

4.1. Methods

4.1.1. Listeners

Twenty-two undergraduate students were recruited from the Department of Psychological and Brain Sciences at the University of Louisville.

4.1.2. Stimuli

The finite-impulse-response filter was modified to have 0-dB gain for frequencies in the range ± 50 Hz around the F_2 of the target vowel and full band rejection at higher and lower frequencies. This filter was applied to the precursor sentences from Experiment 1a and the output was used to estimate the root-mean-square (RMS) amplitude within the passband. A pure tone was then generated at one of the five F_2 center frequencies (1000, 1300, 1600,

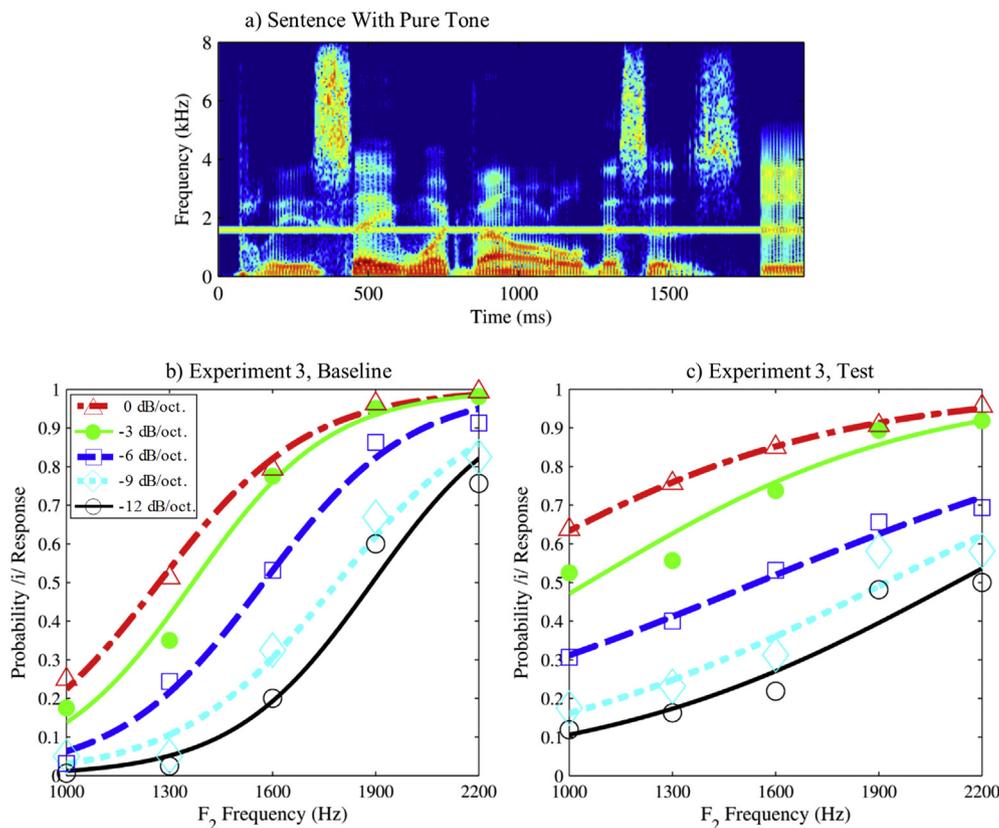


Fig. 4. (a) Spectrogram for a sample trial from Experiment 3. A pure tone was added to the unfiltered sentence at 1600 Hz, the center frequency of F_2 in the following target vowel. (b–c) Group psychometric functions from Experiment 3, following the layout of previous figures. Logistic regressions were fitted to the group data at each level of spectral tilt for illustration purposes. (b) Mean responses to isolated vowels in Experiment 3. (c) Mean responses to vowels following the sentence-plus-tone precursor in Experiment 3.

Table 3

Results of Experiment 3. Mean weights (standardized logistic regression coefficients; Baseline, Test columns) and mean weight changes (Calibration columns) are presented with one standard error of the mean indicated in parentheses.

Experiment	Baseline		Test		Calibration	
	F ₂	Tilt	F ₂	Tilt	ΔF ₂	ΔTilt
3 (<i>n</i> = 19)	1.81 (0.19)	0.85 (0.16)	1.02 (0.20)	1.25 (0.22)	-0.79 (0.17)	+0.40 (0.14)

1900, 2200 Hz) at 1809 ms duration. The RMS amplitude of each tone was adjusted so that when it was added to an unprocessed precursor sentence, the total RMS amplitude in this frequency region equaled that for the filtered sentence (with the added stable spectral peak) as used in Experiment 1a (Fig. 4a). This process was repeated for each F₂ center frequency, creating five unique precursor sentences.

The isolated vowels from Experiment 1a were used in Experiment 3. Target vowels and precursor sentences were again each set to equal RMS amplitude. Precursor sentences and matching target vowels were concatenated with a 50-ms ISI, but the tone continued through the ISI to facilitate perceptual grouping between the tone and the F₂ of the target vowel.

4.1.3. Procedure

Experiment 3 followed the same procedures as the other experiments. The entire session lasted approximately 20 min.

4.2. Results

Two listeners' results were identified as outliers according to Wilcox's (2005) Minimum Generalized Variance method, so their data sets were excluded from further analysis. One additional listener's results were removed due to failure to follow instructions. Psychometric functions for group data across the remaining 19 listeners are presented in Fig. 4. Mean cue weights for these listeners are presented in Table 3. Listeners again weighed F₂ more heavily for vowels in isolation (paired-samples *t*-test: $t_{18} = 2.87$, $p < 0.025$). The weights were similar to those for Experiment 1a (cf. Table 1). Tilt test weights were highly similar to F₂ test weights.

Spectral calibration again occurred in the predicted directions (F₂ weight decrease, one-tailed *t*-test against zero: $t_{18} = 4.70$, $p < 0.001$; tilt weight increase tested against zero: $t_{18} = 2.86$,

$p < 0.01$). The F₂ weight change was modestly but not significantly larger for speech-plus-tone stimuli than for the nonreverberant stimuli tested in Experiment 1a (one-tailed independent-samples *t*-test: $t_{34} = 1.57$, $p = 0.06$). Critically, results of two-tailed independent-samples *t*-tests indicate that the F₂ weight change in Experiment 3 did not differ significantly from those in Experiments 1b and 2, when reverberation was present either with ($t_{33} = 0.39$, $p = 0.70$) or without reverberant tails ($t_{37} = 0.44$, $p = 0.66$). Changes in spectral tilt weights in Experiment 3 were similar to those observed in Experiment 1a with nonreverberant stimuli ($t_{34} = 0.26$, $p = 0.79$) and significantly smaller than the weight changes observed for reverberant stimuli (Experiment 1b: $t_{33} = 2.36$, $p < 0.05$; Experiment 2: $t_{37} = 5.37$, $p < 0.001$).

4.3. Discussion

The results illustrate how listeners compensate for stable spectrotemporal properties during speech perception. A stable spectral peak was added to the unprocessed precursor sentence using a pure tone at the target vowel's F₂ center frequency rather than amplifying that frequency region. This created a spectral peak that was constant across time rather than fluctuating (compare Figs. 4a to 1a). Spectral calibration to F₂ for these nonreverberant speech-plus-tone stimuli was similar to that for reverberant stimuli in Experiments 1b and 2 (one-way ANOVA: $F_{2,52} = 0.39$, $p = 0.68$; Fig. 5c). When these results were pooled into one group and tested against results from Experiment 1a, they revealed significantly larger decreases in F₂ weights for the continuous spectral peaks than for the fluctuating peaks (one-tailed independent-samples *t*-test: $t_{70} = 2.12$, $p < 0.025$). Thus, enhanced spectral calibration was not specific to reverberation.

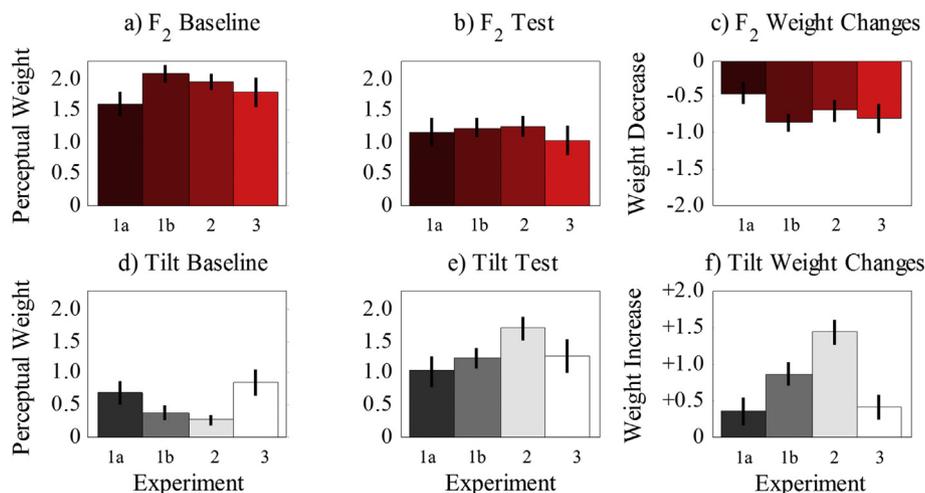


Fig. 5. Results from all experiments arranged by spectral cue. (a) Mean F₂ weights for identifying vowels presented in isolation. (b) Mean F₂ weights for identifying vowels following filtered precursor sentences or the sentence-plus-tone precursor. (c) Mean weight changes (i.e., spectral calibration) to F₂. (d) Mean spectral tilt weights for identifying vowels presented in isolation. (e) Mean spectral tilt weights for identifying vowels following filtered precursor sentences or the sentence-plus-tone precursor sentence. (f) Mean weight changes (i.e., spectral calibration) for spectral tilt. Error bars represent ± 1 standard error of the mean.

5. General discussion

The present experiments investigated how speech perception adjusts to stable spectrotemporal properties of the listening context. Listeners increased the degree to which they compensated for these stable spectral peaks in reverberant stimuli compared to nonreverberant stimuli, as evidenced by larger changes in their cue weights (Experiments 1a and 1b). This perceptual compensation was not due to the presence of the reverberant tails, as removing these tails from the target vowels did not alter listeners' weights for F_2 (Experiment 2). The compensation was revealed not to be due to reverberation at all, as increased spectral calibration to the stable spectral peak was observed when the unprocessed precursor sentence was combined with a pure tone that matched the F_2 of the target vowel (Experiment 3). In all, vowel identification was highly sensitive to stable spectral and spectrotemporal properties of the listening environment without these properties being specific to speech or reverberation *per se*. Low-level signal properties yielded these adjustments in vowel perception irrespective of their source (speech versus tone) or cause (room acoustics versus concurrent tone), suggesting that perceptual compensation for stable spectral properties in the listening environment is a rather general process.

Whether hearing reverberant, filtered, or speech-plus-tone stimuli, listeners exhibited the same general patterns of performance: decreased reliance on the stable spectral property (F_2 ; Fig. 5c) and increased reliance on the changing (and thus informative) spectral property (tilt; 5f) when identifying target vowels. Importantly, the degree to which listeners adjusted their spectral cue weights differed across studies. Listeners decreased F_2 weights by greater degrees to compensate for stable spectral peaks independent of whether this was due to reverberation or constant presentation of a pure tone (5c). However, listeners exhibited larger increases in spectral tilt weights for reverberant stimuli than for nonreverberant stimuli (5f), questioning whether spectral tilt use was also independent of reverberation. The tilt test weights were similar across reverberant and nonreverberant stimuli, except for the unusually high tilt weight observed in Experiment 2 (5e). Baseline weights, however, revealed less reliance on tilt information for reverberant vowels than for nonreverberant vowels (independent-samples *t*-test comparing baseline tilt weights for nonreverberant [Expts. 1a, 3] versus reverberant [Expts. 1b, 2] vowels: $t_{70} = 3.38$, $p < 0.001$) (5d). Since spectral calibration was computed as the difference between baseline and test weights, these smaller baseline weights contributed heavily to larger tilt increases for identifying reverberant vowels. Thus, while decreased reliance on F_2 information was independent of reverberation, it is unclear whether increased reliance on tilt information followed suit.

Calibration to a stable spectral property appears to be independent of the increased perceptual weighting of a changing spectral property. Weight changes for stable and changing spectral cues are typically in opposite directions (decrease in weight for the stable cue; increase in weight for the changing cue) but may not be of equal absolute magnitude (also see Alexander and Kluender, 2010; Stilp and Anderson, 2014). Additionally, these weight changes can be identified separately in the psychometric functions. When comparing test performance to baseline performance in Figs. 2–4, decreased reliance on F_2 corresponds to decreased (shallower) function slopes, whereas increased reliance on spectral tilt corresponds to more positive intercepts (*i.e.*, leftward shifts of psychometric functions).

Perceptual compensation for stable spectrotemporal properties has been demonstrated across a wide variety of tasks using a range of metrics. Here, calibration to stable spectral peaks was measured through changes in cue weights in a vowel identification task (*/i/-*

u/; see also Kiefte and Kluender, 2008; Alexander and Kluender, 2010; Stilp and Anderson, 2014). Darwin et al. (1989) operationalized this compensation using shifts in category boundaries between the vowels */i/* and */ε/*. A similar metric was used to measure compensation for reverberation, measuring phoneme category boundary shifts in a word classification task (along a “*sir*”–“*stir*” continuum; Watkins, 2005a, 2005b; Watkins and Makin, 2007; Watkins et al., 2011). Compensation for reverberation has also been measured as changes in speech reception thresholds for understanding sentences in noise (Brandewie and Zahorik, 2010, 2013; Srinivasan and Zahorik, 2013, 2014).

Results from Experiment 3 are particularly striking because the manipulation of the precursor sentence had little to do with speech. Combining a pure tone with a sentence created a stimulus with two clearly different acoustic sources. Experiments in auditory source segregation often investigate listeners' ability to separate concurrent speech streams (Cherry, 1953; Assmann and Summerfield, 1989) or concurrent tone streams (Miller and Heise, 1950; van Noorden, 1975; Bregman, 1990), but seldom interrogate listeners' ability to segregate a tone from speech (Roberts and Moore, 1991). Stimuli in Experiment 3 gave listeners many ways to segregate these disparate sources: differences in pitch (1000–2200 Hz for the tone, a mean fundamental frequency of 97 Hz for the sentence), timbre, offset (the tone continuing through the 50-ms ISI before the target vowel was presented), patterns of amplitude modulation and frequency modulation (of which the tone had none), and acoustic complexity, among others. It is likely that if listeners were asked to report the number of acoustic streams heard, they would report hearing two. Had listeners ignored the (seemingly irrelevant) tone, smaller calibration might be expected, but this was not observed. Instead, the results are consistent with listeners responding to the speech/tone mixture (Roberts and Moore, 1991). This suggests that the number of perceived streams is irrelevant when compensating for stable spectrotemporal properties of the acoustic environment. Precursors comprised of sentences, a sentence-tone mixture, time-reversed sentences (Kiefte and Kluender, 2008), and even a sawtooth wave (Alexander and Kluender, 2010) all elicit spectral calibration, suggesting that this process depends on the overall statistics of the listening environment. Thus, compensation for stable spectral peaks may be quite general in everyday listening conditions.

What does it mean for a spectral peak in the acoustic environment to be “stable”? Several studies made frequency regions stable through narrowband amplification to introduce a +20 dB peak in the long-term average spectrum (Kiefte and Kluender, 2008; Alexander and Kluender, 2010). One might argue that a frequency region is stable because it has increased energy, as this would promote adaptation of neural responses to frequencies at/near the stable spectral peak and F_2 of the following vowel, diminishing its effectiveness as a speech cue. Simulations of room reverberation are consistent with this idea, as temporal elongation of spectral peaks produced increased energy in the frequency region matching F_2 in the target vowel, and calibration was increased relative to that observed for nonreverberant stimuli. However, two points challenge this idea. First, frequency regions in Experiment 3 (tone added to speech) were configured to match energy in the same frequency regions as for the filtered nonreverberant stimuli in Experiment 1a. Despite similar amounts of acoustic energy in key frequency regions, F_2 calibration was larger in Experiment 3. Second, Stilp and Anderson (2014) reported similar F_2 weight changes when stable spectral peaks in the preceding sentence had levels of +5 dB, +10 dB, or +15 dB. Similar amounts of energy can produce different degrees of calibration (Experiments 1a, 3) while different amounts of energy can produce similar degrees of calibration (Stilp and Anderson, 2014). While increased energy in a given frequency

region is a prerequisite for inducing spectral calibration, the exact amount of energy required is not clear.

The temporal characteristics of spectral peaks may also affect their “stability”. Creation of a stable spectral peak through narrowband amplification adds a peak to the long-term average spectrum, but does not change its modulation rate. Hearing the stable spectral peak more often in a fixed amount of time increases the listener’s opportunities to calibrate to this property. The results from Experiments 1b, 2, and 3 are all consistent with this notion, as increasing the overall duration of the spectral peak (due to temporal smearing from reverberation or constant presentation via a pure tone) contributed to greater calibration. Thus, increased energy and increased occurrence across time appear to jointly determine the stability of a spectral peak in the acoustic environment.

While spectral calibration and compensation for reverberation are both ways that speech perception compensates for stable properties of the listening environment, it is important to acknowledge just how different these processes are. The properties of the acoustic signal responsible for these effects differ widely. Spectral calibration studies are designed to interrogate listeners’ sensitivity to stable energy in a particular frequency region. This signal property is often produced through amplification via filtering, and the resulting spectral peak waxes and wanes with the inherent modulation structure of speech (roughly 3–8 Hz; [Houtgast and Steeneken, 1985](#); [Elliott and Theunissen, 2009](#)) with considerable modulation depth. Adding a tone to an unfiltered sentence in Experiment 3 produced a frequency band with minimal amplitude modulations and limited modulation depth. Nevertheless, spectral calibration to F_2 was enhanced relative to that observed in Experiment 1a, suggesting that amplitude modulations are not critical for spectral calibration to occur.

Compensation for reverberation appears to be based on the amplitude envelopes of frequency bands in the reverberant signal. [Watkins et al. \(2011\)](#) reported similar compensation for reverberation using spectrally intact and noise-vocoded speech signals, which differ widely in temporal fine structure but share similar amplitude envelopes. [Srinivasan and Zahorik \(2014\)](#) used chimeric stimuli where only the amplitude envelope or temporal fine structure of a speech signal was processed in reverberation before these components were recombined. Speech intelligibility improved following experience with the reverberant-envelope signal (consistent with [Brandewie and Zahorik, 2010](#)), but experience with the reverberant-fine structure signal provided no such benefit. This is consistent with the demonstration that prior listening to a particular room improved the detection of amplitude modulation ([Zahorik and Anderson, 2013](#)). Multiple mechanisms might provide perceptual constancy in the presence of reverberation, given that monaural ([Watkins, 2005a, 2005b](#); [Watkins and Makin, 2007](#); [Watkins et al., 2011](#)) and binaural effects have been reported ([Brandewie and Zahorik, 2010, 2013](#); [Srinivasan and Zahorik, 2013, 2014](#)), but both appear to operate on temporal envelope information.

Spectral calibration and compensation for reverberation also appear to differ in terms of where they occur in the auditory system. While the neural locus of spectral calibration remains to be established, [Alexander and Kluender \(2010\)](#) cite the medial olivocochlear reflex (MOCR) as a likely source. Efferent projections from the medial olivary complex to outer hair cells can modulate cochlear gain ([Warr and Guinan, 1979](#)). When acoustic stimulation occurs at the same frequency as, but before, the target signal, cochlear gain at the target signal frequency is reduced. This has been demonstrated in behavioral and computational investigations of temporal effects in masking ([Strickland, 2004, 2008](#); [Jennings et al., 2011](#)), and might explain reduced responsiveness to F_2 in

the target vowel when it follows earlier stimulation at that same frequency (i.e., the stable spectral peak in the precursor sentence).

Recent physiological findings suggest compensation for reverberation in the inferior colliculus (IC). [Slama and Delgutte \(2015\)](#) reported that some IC neurons featured superior coding of reverberant stimuli over anechoic stimuli with matched modulation depths. This is similar to larger neural modulation gains reported for reverberant stimuli than for anechoic stimuli ([Kuwada et al., 2014](#)). In addition, [Slama and Delgutte \(2015\)](#) reported smaller degradation of temporal coding in IC responses than would be suggested by the attenuated modulation depth in the acoustic stimuli. Substrates of compensation for reverberation may occur earlier than the inferior colliculus. [Wickesberg and Oertel \(1990\)](#) reported that inhibitory feedback from the dorsal cochlear nucleus to the ventral cochlear nucleus can produce monaural echo suppression, and aspects of echo suppression are conceivably related to compensation for reverberation (see [Brandewie and Zahorik, 2010](#) for discussion).

6. Summary

While differing in acoustic and likely physiological sources, spectral calibration and compensation for reverberation are two of the many ways that speech perception compensates for regularities in the acoustic environment. Here, stable spectral peaks in reverberant speech and a nonreverberant speech/tone mixture resulted in increased perceptual calibration compared to nonreverberant speech. Given the regularity with which listening environments alter sounds’ spectra and/or introduce reverberation, these compensatory processes likely play a considerable role in everyday speech perception.

Acknowledgements

The authors thank Associate Editor Brian C. J. Moore and two anonymous reviewers for very helpful comments on a previous version of this manuscript. The authors also thank Alexandria Beason, Kara Hendrix, Almira Klanko, Orlando Madriz, Andrew McPheron, Asim Mohiuddin, Emily Nations, Elizabeth Niehaus, Ijeoma Okorie, McKenzie Sexton, and Caitlyn Stromatt for assistance with data collection.

References

- [Alexander, J.M., Kluender, K.R., 2010.](#) Temporal properties of perceptual calibration to local and broad spectral characteristics of a listening context. *J. Acoust. Soc. Am.* 128 (6), 3597–3513.
- [Allen, J.B., Berkley, D.A., 1979.](#) Image model of efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.* 65 (4), 943–950.
- [Assmann, P.F., Summerfield, Q., 1989.](#) Modeling the perception of concurrent vowels: vowels with the same fundamental frequency. *J. Acoust. Soc. Am.* 85 (1), 327–338.
- [Assmann, P.F., Summerfield, Q., 2004.](#) The perception of speech under adverse conditions. In: Greenberg, S., Ainsworth, W.A., Popper, A.N., Fay, R.R. (Eds.), *Speech Processing in the Auditory System*. Springer, New York, pp. 231–308.
- [Attneave, F., 1954.](#) Some informational aspects of visual perception. *Psychol. Rev.* 61 (3), 183–193.
- [Barlow, H.B., 1961.](#) Possible principles underlying the transformations of sensory messages. In: Rosenblith, W.A. (Ed.), *Sensory Communication*. MIT Press, Cambridge, pp. 53–85 (Mass. and John Wiley, New York).
- [Beeston, A.V., Brown, G.J., Watkins, A.J., 2014.](#) Perceptual compensation for the effects of reverberation on consonant identification: evidence from studies with monaural stimuli. *J. Acoust. Soc. Am.* 136 (6), 3072–3084.
- [Beutnagel, M., Conkie, A., Schroeter, J., Stylianou, Y., Syrdal, A., 1997.](#) AT&T Natural Voices Text-to-speech [Computer Software]. Retrieved September 4, 2014, from <http://www.research.att.com/~ttsweb/tts/demo.php>.
- [Brandewie, E.J., Zahorik, P., 2010.](#) Prior listening in rooms improves speech intelligibility. *J. Acoust. Soc. Am.* 128 (1), 291–299.
- [Brandewie, E., Zahorik, P., 2013.](#) Time course of a perceptual enhancement effect for noise-masked speech in reverberant environments. *J. Acoust. Soc. Am.* 134 (2), EL265–EL270.

- Bregman, A.S., 1990. *Auditory Scene Analysis: the Perceptual Organization of Sound*. MIT Press, Cambridge, MA.
- Cherry, E.C., 1953. Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25 (5), 975–979.
- Darwin, C.J., McKeown, J.D., Kirby, D., 1989. Perceptual compensation for transmission channel and speaker effects on vowel quality. *Speech Commun.* 8 (3), 221–234.
- Elliott, T.M., Theunissen, F.E., 2009. The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* 5 (3), e1000302.
- Houtgast, T., Steeneken, H.J., 1973. The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acta Acust. United Acust.* 28 (1), 66–73.
- Houtgast, T., Steeneken, H.J., 1985. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* 77 (3), 1069–1077.
- ISO-3382, 1997. *Acoustics – Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters*. International Organization of Standardization, Geneva.
- Jennings, S.G., Heinz, M.G., Strickland, E.A., 2011. Evaluating adaptation and olivocochlear efferent feedback as potential explanations of psychophysical overshoot. *J. Assoc. Res. Otolaryngol.* 12 (3), 345–360.
- Kiefte, M., Kluender, K.R., 2008. Absorption of reliable spectral characteristics in auditory perception. *J. Acoust. Soc. Am.* 123 (1), 366–376.
- Klatt, D.H., Klatt, L.C., 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 87 (2), 820–857.
- Knudsen, V.O., 1929. The hearing of speech in auditoriums. *J. Acoust. Soc. Am.* 1 (1), 30.
- Kuwada, S., Bishop, B., Kim, D.O., 2014. Azimuth and envelope coding in the inferior colliculus of the unanesthetized rabbit: effect of reverberation and distance. *J. Neurophysiol.* 112, 1340–1355.
- Miller, G.A., Heise, G.A., 1950. The trill threshold. *J. Acoust. Soc. Am.* 22, 637–638.
- Nábělek, A.K., Dagenais, P.A., 1986. Vowel errors in noise and in reverberation by hearing-impaired listeners. *J. Acoust. Soc. Am.* 80 (3), 741–748.
- Nábělek, A.K., Donahue, A.M., 1984. Perception of consonants in reverberation by native and non-native listeners. *J. Acoust. Soc. Am.* 75 (2), 632–634.
- Nábělek, A.K., Letowski, T.R., 1985. Vowel confusions of hearing-impaired listeners under reverberant and nonreverberant conditions. *J. Speech Hear. Disord.* 50 (2), 126–131.
- Nábělek, A.K., Robinson, P.K., 1982. Monaural and binaural speech perception in reverberation for listeners of various ages. *J. Acoust. Soc. Am.* 71 (5), 1242–1248.
- Roberts, B., Moore, B.C.J., 1991. The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony. *J. Acoust. Soc. Am.* 89, 2922–2932.
- Slama, M.C., Delgutte, B., 2015. Neural coding of sound envelope in reverberant environments. *J. Neurosci.* 35 (10), 4452–4468.
- Srinivasan, N.K., Zahorik, P., 2013. Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences. *J. Acoust. Soc. Am.* 133 (1), EL33–EL39.
- Srinivasan, N.K., Zahorik, P., 2014. Enhancement of speech intelligibility in reverberant rooms: role of amplitude envelope and temporal fine structure. *J. Acoust. Soc. Am.* 135 (6), EL239–EL245.
- Stilp, C.E., Alexander, J.M., Kiefte, M., Kluender, K.R., 2010. Auditory color constancy: calibration to reliable spectral properties across nonspeech context and targets. *Atten. Percept. Psychophys.* 72 (2), 470–480.
- Stilp, C.E., Anderson, P.W., 2014. Modest, reliable spectral peaks in preceding sounds influence vowel perception. *J. Acoust. Soc. Am.* 136 (5), EL383–EL389.
- Strickland, E.A., 2004. The temporal effect with notched-noise maskers: analysis in terms of input–output functions. *J. Acoust. Soc. Am.* 115 (5), 2234–2245.
- Strickland, E.A., 2008. The relationship between precursor level and the temporal effect. *J. Acoust. Soc. Am.* 123 (2), 946–954.
- van Noorden, L.P.A.S., 1975. *Temporal Coherence in the Perception of Tone Sequences*. Doctoral dissertation. Eindhoven University of Technology, Netherlands.
- Warr, W.B., Guinan, J.J., 1979. Efferent innervation of the organ of Corti: two separate systems. *Brain Res.* 173 (1), 152–155.
- Watkins, A.J., 2005a. Perceptual compensation for effects of reverberation in speech identification. *J. Acoust. Soc. Am.* 118 (1), 249–262.
- Watkins, A.J., 2005b. Perceptual compensation for effects of echo and of reverberation on speech identification. *Acta Acust. United Acust.* 91 (5), 892–901.
- Watkins, A.J., Makin, S.J., 2007. Steady-spectrum contexts and perceptual compensation for reverberation in speech identification. *J. Acoust. Soc. Am.* 121 (1), 257–266.
- Watkins, A.J., Raimond, A.P., 2013. Perceptual compensation when isolated test words are heard in room reverberation. In: Moore, Brian C.J., Patterson, Roy D., Winter, Ian M., Carlyon, Robert P., Gockel, Hedwig E. (Eds.), *Basic Aspects of Hearing*. Springer, New York, pp. 193–201.
- Watkins, A.J., Raimond, A.P., Makin, S.J., 2011. Temporal-envelope constancy of speech in rooms and the perceptual weighting of frequency bands. *J. Acoust. Soc. Am.* 130 (5), 2777–2788.
- Wickesberg, R.E., Oertel, D., 1990. Delayed, frequency-specific inhibition in the cochlear nuclei of mice: a mechanism for monaural echo suppression. *J. Neurosci.* 10 (6), 1762–1768.
- Wilcox, R.R., 2005. *Introduction to Robust Estimation and Hypothesis Testing*, second ed. Elsevier, London.
- Winn, M.B., Litovsky, R.Y., 2015. Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *J. Acoust. Soc. Am.* 137 (3), 1430–1442.
- Zahorik, P., 2009. Perceptually relevant parameters for virtual listening simulation of small room acoustics. *J. Acoust. Soc. Am.* 126 (2), 776–791.
- Zahorik, P., Anderson, P.W., 2013. Amplitude modulation detection by human listeners in reverberant sound fields: effects of prior listening exposure. *Proc. Meet. Acoust.* 19, 050139.