Attention, Perception, & Psychophysics
2010, 72 (2), 470-480
doi:10.3758/APP.72.2.470

Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and targets

Christian E. Stilp
University of Wisconsin, Madison, Wisconsin

Joshua M. Alexander
Purdue University, West Lafayette, Indiana

Michael Kiefte
Dalhousie University, Halifax, Nova Scotia, Canada

and

Keith R. Kluender
University of Wisconsin, Madison, Wisconsin

Brief experience with reliable spectral characteristics of a listening context can markedly alter perception of subsequent speech sounds, and parallels have been drawn between auditory compensation for listening context and visual color constancy. In order to better evaluate such an analogy, the generality of acoustic context effects for sounds with spectral–temporal compositions distinct from speech was investigated. Listeners identified nonspeech sounds—extensively edited samples produced by a French horn and a tenor saxophone—following either resynthesized speech or a short passage of music. Preceding contexts were "colored" by spectral envelope difference filters, which were created to emphasize differences between French horn and saxophone spectra. Listeners were more likely to report hearing a saxophone when the stimulus followed a context filtered to emphasize spectral characteristics of the French horn, and vice versa. Despite clear changes in apparent acoustic source, the auditory system calibrated to relatively predictable spectral characteristics of filtered context, differentially affecting perception of subsequent target nonspeech sounds. This calibration to listening context and relative indifference to acoustic sources operates much like visual color constancy, for which reliable properties of the spectrum of illumination are factored out of perception of color.

To be effective, sensorineural systems must maintain perceptual stability across substantial energy flux in the environment. In vision, intensity and spectral composition of reflected light entering the eye vary dramatically depending on illumination, yet viewers perceive objects as having relatively constant brightness and color. Spectral distribution of light entering the eye depends on both the spectrum of illumination and the spectral characteristics that light encounters on its path to the eye (Nassau, 1983). In order to achieve color constancy, the visual system must extract reliable spectral properties across the entire image in order to determine inherent spectral properties of objects within the scene (Boynton, 1988; Churchland & Sejnowski, 1988; Foster et al., 1997).

At least two kinds of visual processes underlie perception of color across changes in the spectrum of illumination (Arend & Reeves, 1986). The first requires the visual system to become accustomed to the illuminant through light (Bramwell & Hurlbert, 1996; von Kries, 1905; Whittle, 1996) and contrast adaptation (Brown & MacLeod, 1997; Webster & Mollon, 1995). A second type of process is more immediate, involving moving the eye rapidly across a scene, sampling the illumination spectrum, and gaining information from light changes accompanying eye movements (Foster, Amano, & Nascimento, 2001; Zaidi, Spehar, & DeBonet, 1997).

For color constancy, one can consider the spectrum of illumination as a filter imposed on the full context of viewing, and perception of constant color is maintained by relative differences between the spectral composition of the object being viewed versus reliable spectral characteristics of the viewing context. In this way, perception is normalized, or calibrated, with respect to the imposing filter common to both context and object.

Multiple experiments have now demonstrated that auditory perception of speech relies on properties of the listen-

C. E. Stilp, cestilp@wisc.edu

ing context in ways quite similar to visual color constancy. In a classic study on context effects in vowel perception, Ladefoged and Broadbent (1957) showed that identification of a target vowel from a synthetic [bɪt] (lower first formant frequency, F_1) to [bɛt] (higher F_1) series was affected by manipulations of average F_1 in a preceding context sentence. Raising average F_1 frequency in the context sentence elicited more /bɪt/ (lower F_1) responses. Ladefoged and Broadbent drew explicit analogies between their findings and color constancy in vision. They wrote

> It is obvious that this experiment provides a demonstration of perceptual constancy in the auditory field; that is an auditory phenomenon somewhat parallel to the visual case in which the response evoked by a stimulus is influenced by the stimuli with which it is closely associated. An example is the correct identification of the color of an object in widely differing illuminations. Consequently it is hoped that further investigation of the auditory phenomenon will provide data which are of general psychological interest. (p. 102)

Although Ladefoged and Broadbent's (1957) experiments may or may not have provided a close analogy to color constancy as they imagined, the relationship between effects of auditory context and color constancy has received relatively little attention over the past half century. However, considerable research has been conducted investigating effects of listening context on perception of speech sounds, and findings are consistent with Ladefoged and Broadbent's speculation. For example, Watkins and Makin (1994) argued that it was not specific F_1 frequencies per se that shifted responses in the studies by Ladefoged and Broadbent, but rather the long-term spectrum of the context sentence. Watkins (1991) demonstrated effects similar to those observed by Ladefoged and Broadbent by filtering a precursor sentence with the difference between spectral envelopes of two vowels. This resulted in a context colored by spectral peaks of one vowel and by spectral notches corresponding to the peaks of the other vowel. When the context sentence was processed by a difference filter with the spectral shape of /ɪ/ minus /ɛ/, there was an increase in the number of /ɛ/ responses to an /ɪtʃ/–/ɛtʃ/ series. This perceptual shift was observed by using many different speech contexts varying in talker gender, spatial position, ear of presentation, whether the context was forward or time-reversed, and even when it was speech-shaped, signal-correlated noise. In each case, perception calibrated to persistent spectral peaks and notches of the context emphasizing /ɪ/, such that listeners were more likely to hear target vowels as /ɛ/.

The extent to which nonspeech acoustic contexts influence perception is less clear. Watkins (1991) processed speech-shaped, signal-correlated noise using /ɪ/ − /ɛ/ spectral envelope difference filters and reported contrast effects on identification of target vowels similar to those observed when the context was speech. He reported smaller but statistically significant perceptual effects from filtered noise contexts in a diotic task; however, these smaller effects could not be replicated with dichotic presentations for which noise contexts and speech targets were perceived to arise from different spatial locations. From these results, Watkins (1991) suggested that speech has a context effect but noise does not, since it originates from a different source than the target. On the other hand, Holt (2005, 2006a, 2006b) demonstrated that preceding sequences of pure tones influence perception of a /da/–/ga/ target series. Consequently, it remains unclear as to whether effects of listening context on perception of speech sounds depend on the preceding acoustic context being speech or speechlike or sharing the same apparent source.

Kiefte and Kluender (2008) recently reported experiments designed to assess relative contributions of spectrally global (spectral tilt) versus local (spectral peak) characteristics of a listening context on identification of vowel sounds. They varied both spectral tilt and center frequency of the second formant (F_2) to generate a matrix of vowel sounds that perceptually varied from /u/ to /i/. Listeners identified these vowels following filtered forward or time-reversed precursor sentences. When precursor sentences were filtered to share the same long-term spectral tilt as the target vowel, tilt information was neglected and listeners identified vowels principally on the basis of F_2. Conversely, when precursors were filtered with a single pole centered at the F_2 frequency of the target vowel, perception instead relied on tilt. These results demonstrate calibration to reliable global and local spectral features across both intelligible and unintelligible speechlike contexts.

Most recently, Alexander and Kluender (2009) created nonspeech precursor contexts consisting of a harmonic spectrum filtered by four frequency-modulated resonances (somewhat akin to formants). Precursors filtered to match F_2 or tilt of following vowels induced perceptual calibration (i.e., diminished perceptual weight) to F_2 and tilt, respectively. Perceptual calibration to F_2 and tilt followed different time courses. Calibration to F_2 was greatest for shorter duration precursors; in contrast, calibration to tilt was greatest for precursors that provided greater opportunity to sample the spectrum (longer duration and/or higher resonance-modulation rates).

Analogous to vision, spectral composition of sound entering the ear is colored by the listening environment. Energy at some frequencies is emphasized by acoustically reflective properties of surfaces, whereas energy at other frequencies is attenuated by acoustic absorbent materials. In this way, listening context spectrally shapes the acoustic signal. Consistent with Ladefoged and Broadbent (1957), Kluender and colleagues (Kluender & Alexander, 2008; Kluender & Kiefte, 2006) have proposed that effects of listening context are, in fact, closely analogous to visual color constancy.

However attractive one finds these parallels between visual color constancy and auditory calibration to reliable characteristics of a listening context, all listening studies discussed above employed speech or speechlike stimuli as targets to be identified. Previous experiments were designed to better understand speech perception or effects of adverse conditions on speech perception. Speech stimuli were appropriate and also expeditious because they pro-

vide information with which the listener has extensive experience, whether played to different ears, from different spatial positions, or conveyed in speech-correlated noise (Watkins, 1991). Even time-reversed speech is readily recognizable as speech because it contains spectral and temporal fluctuations comparable to forward samples. Listeners may perceive targets and contexts relative to expectations given their extensive experience with similar stimuli. As such, fundamental processes responsible for calibration to listening context are conflated with extensive knowledge and experience gained through listening to speech in multiple acoustic environments. For example, spectral modifications to acoustic contexts may be perceptually salient by virtue of comparison with the listener's experience.

By contrast, most studies of visual color constancy do not use familiar surfaces. Visual color constancy does not require any familiarity with surfaces or objects, as color constancy maintains to unfamiliar color patches, such as Mondrians (e.g., Land, 1983; McCann, McKee, & Taylor, 1976). Despite speech sounds being highly familiar and plentiful in the human ecology, they are but a subset of environmental sounds. To better understand auditory constancy across listening contexts and acoustic events, sounds other than speech must be tested as both contexts and as targets. The extent to which such auditory constancy effects extend to nonspeech targets has been unclear, rendering the analogy to visual color constancy potentially premature.

To date, there exists one published letter reporting effects of acoustic context on perception of putatively nonspeech sounds.[1] Stephens and Holt (2003) assessed listeners' discrimination of F_2 and F_3 transitions excised from a /da/–/ga/ series following a speech context of either /al/ or /ar/. Most listeners were unable to consistently classify these speech fragments as either /d/ or /g/ and classification results otherwise differed significantly from those for full /da/–/ga/ syllables. Listeners' discrimination of /d,g/ fragments did, however, vary significantly as a function of preceding speech context.

Data comparing context effects for full-syllable speech and fragments of speech are suggestive, but inconclusive, concerning the claim that the same auditory processes are at work for both speech and nonspeech sounds. First, there is the argument that convergences between speech and nonspeech conditions exist because "nonspeech stimuli that are sufficiently speech-like are processed by central processes involved in speech perception, even when the listener is not aware of the speech-likeness of the stimuli" (Pisoni, 1987, p. 266). Fragments of /d/ and /g/ may fit this definition of sufficient likeness to full /d/ and /g/.

Perception of even full-syllable /da/ and /ga/, however, does not pose a particularly stringent test of context effects. Discrimination of /d/ versus /g/ is not perceptually robust, in that perceptual confusions between naturally spoken /d/ and /g/ are among the most common (Miller & Nicely, 1955). Consequently, perception of synthetic /d/ versus /g/ signaled solely by changes in F_3 transitions is especially labile and, as Stephens and Holt's (2003) data attest, classification of F_3 fragments alone is even less reliable.

Making an auditory analogy to visual color constancy requires investigation of context effects on target sounds that are not speech, but are nonetheless classified fairly consistently absent experimental effects of context. Speech is not the only class of sound that exhibits spectral characteristics important for classification. For example, spectral shape is a critical element in the timbre of musical instruments and consistently has been shown to be among the primary dimensions that listeners use in instrument-classification tasks (e.g., Grey, 1975; Krumhansl, 1989; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995). The extent to which identification of musical instruments, or of any other nonspeech sounds, is calibrated to reliable characteristics of a listening context is unknown. Comparable results across studies employing speech and nonspeech target sounds would strengthen the claim that general auditory processes underlie calibration to the listening environment in support of auditory color constancy. Furthermore, although not entirely novel to listeners, edited instrument samples serve as effective nonspeech sounds because their familiarity to listeners pales in comparison with extensive experience with speech. Should effects of listening context not affect perception of nonspeech sounds, the analogy to visual color constancy does not maintain or, at least, must be modified to apply only to highly familiar sounds such as speech.

The present experiments measure the extent to which characteristics of listening context influence perception of nonspeech stimuli—specifically, musical sounds. Nonspeech contexts and targets are used here to investigate calibration to reliable spectral characteristics across different sources and spectral compositions. Experiment 1 investigates perception of musical instrument targets following a speech context filtered to share the spectral shape of one instrument and the inverse spectral shape of another (following Watkins, 1991; Watkins & Makin, 1994). Aside from the readily apparent change in acoustic source (speech context to music target), Experiment 1 closely follows the design of all earlier examinations of listening context effects by employing speech, a stimulus with which listeners have incomparable experience. Experiment 2 extends these findings by employing an acoustic context consisting of an unrelated musical passage, thereby examining whether perceptual constancy maintains across both contexts and targets with which listeners have comparatively little experience. The hypothesis at test is whether, like the visual system, the auditory system automatically calibrates to the perceptual context, independent of familiarity and apparent source. It is predicted that perception of target sounds will be affected differentially by filtered contexts, such that contexts filtered to emphasize spectral characteristics of a French horn will elicit more "saxophone" identifications, and vice versa.

EXPERIMENT 1

Experiment 1 was designed to investigate effects of filtered speech contexts on perception of unrelated musical instrument sounds. Listeners were asked to classify members of a series of six musical instrument sounds that

varied spectrally from a French horn to a tenor saxophone. The target sound followed a brief sentence context that had either been filtered or served as an unfiltered control. Listeners were asked to classify the target as French horn or saxophone.

Method

Participants. Twenty-five undergraduate students (18–22 years of age) were recruited from the Department of Psychology at the University of Wisconsin, Madison. Consistent with the demographics of the university's psychology majors, approximately two thirds of participants were women. No participant reported any hearing impairment, and all received course credit for their participation.

Stimuli. Base materials for musical instrument stimuli were selected from the McGill University Musical Samples database (Opolko & Wapnick, 1989). Samples of a tenor saxophone and a French horn, each playing the note G3 (196 Hz) and sampled at 44.1 kHz, were selected on the basis of having relatively distinct and harmonically rich spectra. Three consecutive pitch pulses (15.31 msec) of constant amplitude were excised at zero crossings from the center of each sample, matched in fundamental frequency (f_0), and iterated to 140-msec total duration in Praat (Boersma & Weenink, 2007). Stimuli were then weighted by 5-msec linear onset/offset ramps and proportionately mixed in six 6-dB steps to form a series in which the amplitude of one instrument was +30, +18, +6, −6, −18, or −30 dB relative to the other (see Figure 1). Composite stimuli were judged by the authors to sound perceptually coherent (i.e., as if they were produced by a single instrument). Stimuli with 30-dB differences between instruments served as series endpoints, which two of the authors (C.E.S. and K.R.K.) judged to be perceptually indistinguishable from pure (French horn or tenor saxophone) waveforms without mixing. Waveforms were then low-pass filtered with 10-kHz cutoff by using a 10th-order, elliptical infinite impulse response (IIR) filter. Instrument mixing and filtering were performed in MATLAB.

The precursor speech context was the phrase, "You will hear" (1.00-sec duration[2]), spoken by C.E.S. (see Figure 3A). The context was recorded in a single-walled soundproof booth (IAC) using an Audio-Technica AE4100 microphone, amplified, and digitized (44.1 kHz, 16-bit, TDT System II) prior to analysis.

Similar to Watkins (1991), endpoint French horn and tenor saxophone stimuli were analyzed to create spectral envelope difference filters. Spectral envelopes for each instrument were derived from 512-point Fourier transforms, which were smoothed using a 256-point Hamming window with 128-point overlap (Figure 2). Spectral envelopes of each instrument (Figures 2A and 2B) were equated for peak power (in decibels), then subtracted from one another. A 100-point finite impulse response was obtained for each spectral envelope difference (French horn − saxophone and saxophone − French horn) via inverse Fourier transform, generating linear phase filters. Filter responses are plotted in Figures 2C and 2D.

The speech context "You will hear" was processed by each spectral envelope difference filter. These two filtered contexts and one unfiltered control context were low-pass filtered with 10-kHz cutoff using the same elliptical IIR filter as was used for the target series (see Figures 3B–3D). All contexts and targets were then RMS matched in amplitude. Each of the six target stimuli was concatenated to each of three contexts (French horn − saxophone filtered, saxophone − French horn filtered, and unfiltered control), making 18 pairings in all. Finally, the two target endpoints, absent any preceding context, were also RMS matched for use in a familiarization task.

Procedure. Contexts and targets were upsampled to 48828 Hz, converted from digital to analog (Tucker–Davis Technology RP2), amplified (TDT HB4), and presented diotically over headphones (Beyer DT 150) at 72 dB_{SPL}. Experiments were conducted in four parts with 1–3 listeners participating concurrently in single-subject soundproof booths. First, participants were familiarized with target endpoints by hearing each instrument series endpoint labeled and played twice. Second, listeners identified target endpoints presented in isolation by pressing buttons labeled "French horn" and "Saxophone" on a response box, without receiving any feedback. Each endpoint stimulus was presented 50 times in random order (100 total responses) in a 5-min session. Third, the 18 context–target sequences were presented five times each in random order during each of two 15-min sessions for a total of 180 responses from every listener. Listeners were given the opportunity to take a short break between these two longer sessions; otherwise, all sessions immediately followed one another. Each listener received a different random stimulus order in each session. Finally, at completion of the experiment, listeners completed a brief questionnaire regarding their musical expertise. The questionnaire asked them to rate their skill level in musical performance on a 1–5 Likert-type scale and to list all experience performing music in solo and ensemble formats. The entire experiment took approximately 40 min.

Results

Listeners were required to meet a performance criterion of at least 90% correct in the familiarization task. Seven failed to meet this criterion, and their data were removed from further analysis. Experiment 1 results are shown in Figure 4A as identification functions. The probability of a "saxophone" response is plotted as a function of target stimulus, with the French horn endpoint denoted as "1" and the tenor saxophone endpoint denoted as "6." Responses in each of the three context conditions are represented by separate lines in the figure (see legend). Preceding speech context differentially altered perception of subsequent musical instrument sounds. Individual listener's responses were fit via logistic regression (McCullagh & Nelder, 1989), and the 50% crossover for each context condition was estimated from regression coefficients (Figure 4B). Crossover points were analyzed in a one-way, repeated measures ANOVA with three levels of context (French horn − saxophone, saxophone − French horn, and unfiltered control) as the sole factor. The main effect of context was significant [$F(1.31,22.27) = 5.44$, $p < .05$, with Greenhouse–Geiser sphericity correction]. Post hoc tests using Tukey's honestly significant difference (HSD) indicated that 50% crossover for targets following the French horn − saxophone filtered context was significantly lower (i.e., a leftward shift of the identification function) than for the saxophone − French horn filtered context ($\alpha = .05$).

Finally, the influence of musical context was independent of participants' musical ability. Questionnaire items (performance skill rating, total years of solo performance experience, and total years of ensemble performance experience) were entered into a linear regression, with the total context effect (the difference in estimated 50% crossovers for saxophone − French horn and French horn − saxophone context conditions) as the dependent measure. If musical experience influenced the size of the context effect, one would predict extensive musical background to contribute to these significant differences in mean boundary locations. No such relationship was observed in the regression ($r^2 = .23$, n.s.). Analyses of alternative nonlinear relationships (logarithmic, quadratic, cubic) between performance and musical experience also yielded no reliable fits to the data.

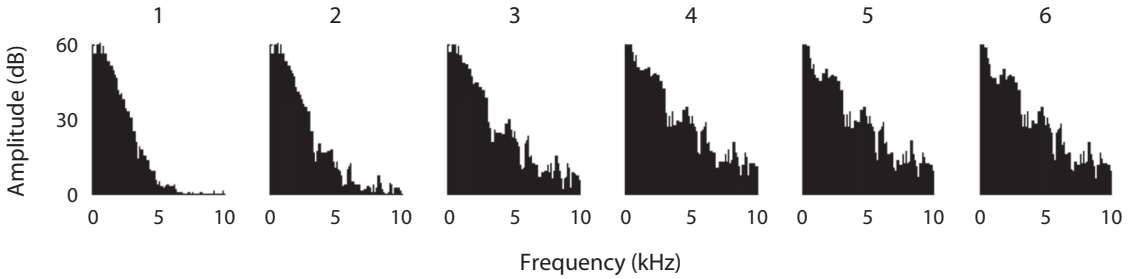

Figure 1. Power spectra denoting instrument series. Spectra vary from French horn endpoint (1, far left; French horn is +30 dB relative to saxophone) to saxophone endpoint (6, far right; French horn is −30 dB relative to saxophone) in 12-dB steps.

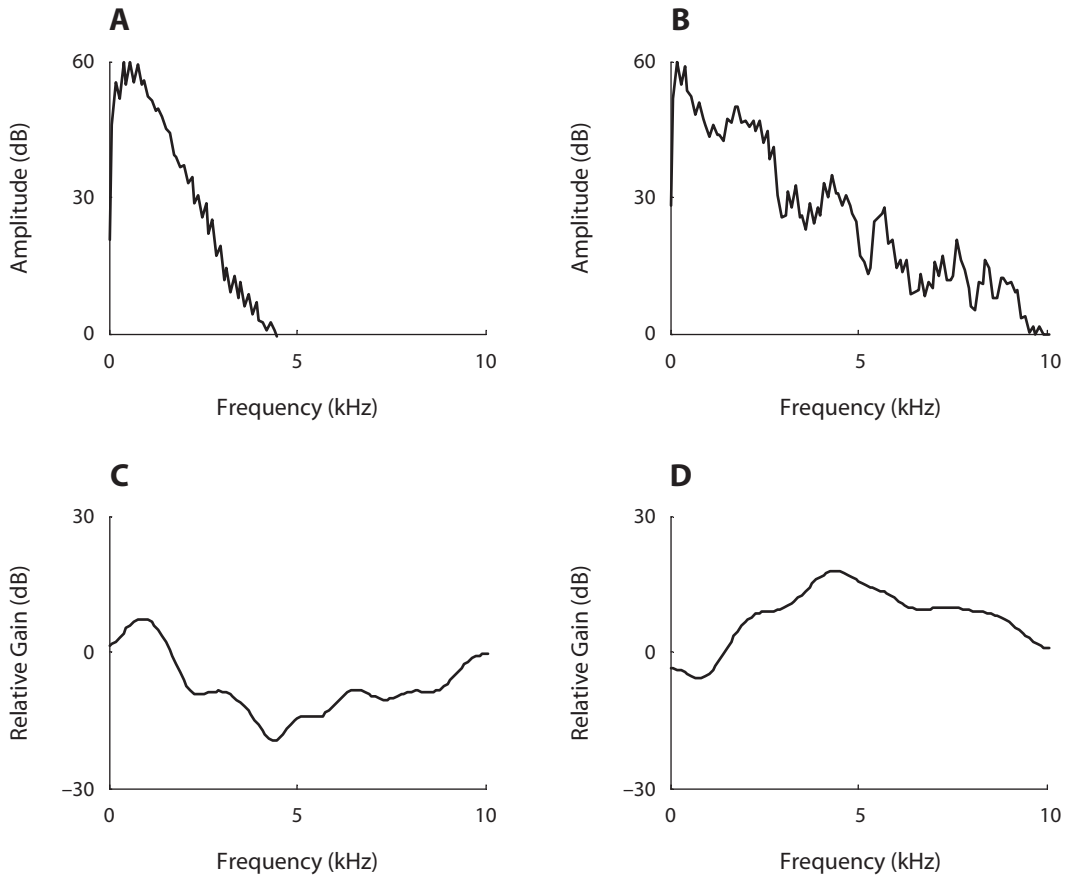

Figure 2. Plots denoting the generation of spectral envelope difference filters. (A) Power spectral envelope of the French horn endpoint. (B) Power spectral envelope of the tenor saxophone endpoint. Envelopes were obtained using 512-point fast Fourier transforms, smoothed with 256-point Hamming windows with 50% overlap. (C) Smoothed spectral envelope difference filter response of the French horn spectral envelope minus the saxophone spectral envelope. With negligible energy in the French horn above 5 kHz, relative gain of the filter above this point is negative. (D) Smoothed difference filter response of the saxophone spectral envelope minus the French horn envelope. The filter response in panel D is approximately the inverse of the filter response shown in panel C.

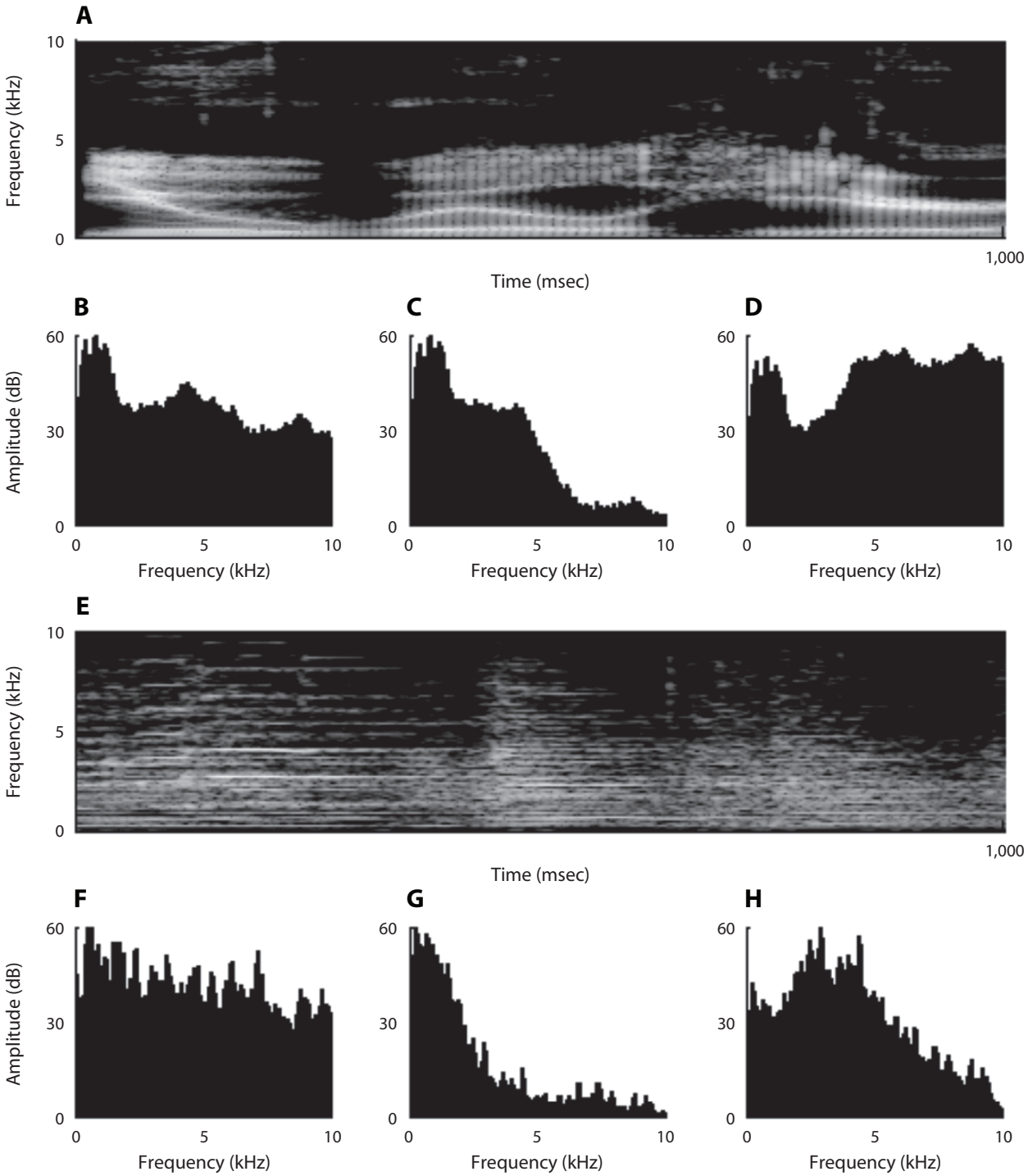

Figure 3. Acoustic contexts and filtering in Experiments 1 and 2. (A) Spectrogram of the speech context "You will hear" presented in Experiment 1. Panels in the second row show power spectral densities of the unfiltered speech context (B), context filtered to emphasize spectral properties of the French horn (C), and context filtered to emphasize spectral properties of the saxophone (D). (E) Spectrogram of the string quintet context presented in Experiment 2. Panels in the fourth row show power spectral densities of the unfiltered music context (F), context filtered to emphasize spectral properties of the French horn (G), and context filtered to emphasize spectral properties of the saxophone (H).

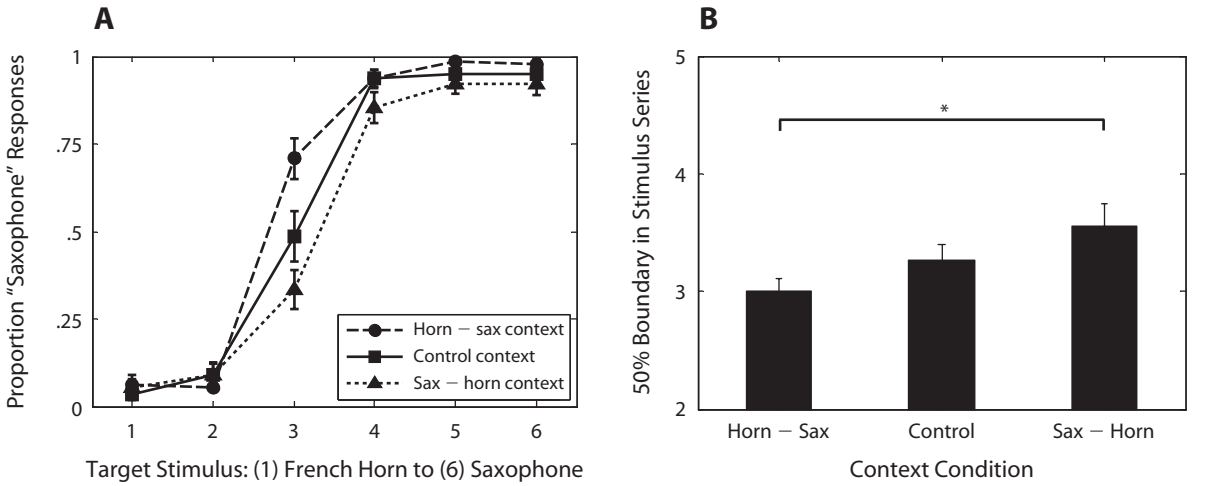

Figure 4. Results from 18 listeners in Experiment 1. (A) Proportion "saxophone" responses as a function of target stimulus. Steps along the abscissa correspond to the series of target sounds, with 1 corresponding to the French horn endpoint and 6 to the tenor saxophone endpoint. Separate lines denote performance in the French horn − saxophone filtered (abbreviated "Horn − Sax"), unfiltered ("Control"), and saxophone − French horn filtered ("Sax − Horn") context conditions. Standard error bars are depicted for each of the 18 trial types. (B) Mean 50% crossovers for each context condition derived from fitted logistic regression functions. The ordinate corresponds to the location in the series of six target stimuli (1 = *French horn endpoint*; 6 = *tenor saxophone endpoint* [neither shown]), where listeners are projected to respond "French horn" and "saxophone" with equal frequency. Error bars denote standard errors of the means. $^{*}p < .05$.

Discussion

Filtered speech contexts influenced perception of musical instrument sounds. This demonstrates that context effects persist despite obvious differences between acoustic sources for context and target. Preceding speech context processed by a French horn − saxophone filter encouraged more "saxophone" responses than the same context processed by the saxophone − French horn filter.[3] Perceptual calibration to predictable spectral characteristics of a listening context appears to be indifferent to a change in sound source. The current findings demonstrate generalization of context effects to nonspeech, musical instrument targets with which listeners have considerably less experience than they do with familiar speech sounds.

All prior examinations of contrast effects in audition, including Experiment 1, have used speech as context, target, or both (e.g., Alexander & Kluender, 2009; Holt, 2005, 2006a, 2006b; Kiefte & Kluender, 2008; Watkins, 1991; Watkins & Makin, 1994). This methodological decision persistently obscures examination of contrast effects independent of extensive experience or knowledge about the signal, such as listeners have with speech. This point served as the authors' motivation for using modified musical instrument samples as nonspeech targets. Experiment 1 demonstrated that listeners need not have extensive experience with target sounds for perceptual calibration to occur. However, preceding speech contexts were used. Some possibility remains that listeners perceived target sounds relative to expectations of speech, given their extensive experience and familiarity with the acoustic context that preceded target sounds. For example, although listeners were comparatively unfamiliar with target sounds, they were highly experienced hearing speech under different listening conditions. Such knowledge of how speech contexts habitually sound may be brought to bear when identifying following targets.

To fully control for effects of experience and familiarity, a second experiment was conducted that employed unfamiliar musical sounds as both context and target. Following the results of Experiment 1, observing spectral contrast following a readily apparent change between two relatively unfamiliar acoustic sources would suggest that the auditory system is largely indifferent to source when calibrating to listening context.

EXPERIMENT 2

If effects of listening context operate similarly for perception of both musical instrument sounds and familiar speech sounds following speech precursors, it remains to be demonstrated whether nonspeech precursor contexts provide the same perceptual effects. In Experiment 2, music was used for both context and target. Both preceding acoustic context and target sounds arise from acoustically and perceptually distinct sources that are both relatively unfamiliar to listeners, especially when compared with the extensive familiarity of speech. Given the results of Experiment 1 together with earlier findings (Holt, 2005, 2006a, 2006b; Kiefte & Kluender, 2008; Watkins, 1991), details of acoustic context relative to target items may be relatively immaterial, as long as the context conveys reliable spectral characteristics.

Method

Participants. Twenty-five native English speakers (18–22 years of age) were recruited from the Department of Psychology at the University of Wisconsin, Madison. None reported any hearing impairment and all received course credit for their participation. No listeners had participated in Experiment 1.

Stimuli. The musical context was an excerpt (1.00 sec) of a musical selection (Franz Schubert's String Quintet in C Major, Allegretto) taken from compact disc (Figure 3E). This selection was chosen because it is both spectrally rich and quite distinct from the saxophone–French horn series, consisting solely of five string instruments playing concurrently. The selection was also expected to be unfamiliar to participants in the study. The music context was processed by both spectral envelope difference filters described in Experiment 1 (see Figures 3F–3H). The same instrument stimuli from Experiment 1 were used as identification targets in Experiment 2. As in Experiment 1, each of the six instrument stimuli was concatenated to the three new contexts, making 18 pairings in all. Each pairing was then upsampled to 48828 Hz and RMS normalized in MATLAB.

Procedure. The experimental procedure was the same as described in Experiment 1, with the addition of one questionnaire item. Listeners were asked whether they had previously heard the musical context or could identify its composer. The entire experiment took approximately 40 min.

Results

Of the 25 listeners recruited for Experiment 2, 10 failed to meet the performance criterion of 90% correct in instrument endpoint labeling absent preceding acoustic context. Therefore, their results were excluded from further analysis. Data from the remaining 15 listeners are shown in Figure 5A as identification functions. Similar to Figure 4A, the probability of a "saxophone" response is plotted as a function of target instrument, with separate lines representing responses in different context conditions (see legend).

As in Experiment 1, perception of the target stimulus was differentially affected by the preceding context. Each listener's data was fit with a logistic regression, and 50% crossovers were estimated from the fitted regression functions (Figure 5B). These crossovers were subjected to a one-way, repeated measures ANOVA with three levels of context as the sole factor. The main effect of musical context was significant [$F(1.41,19.72) = 8.31$, $p < .01$, with Greenhouse–Geiser sphericity correction]. Tukey HSD tests indicate that targets following the context processed by the French horn − saxophone filter elicited more "saxophone" responses than did either the control condition or the saxophone − French horn filtered context ($\alpha = .05$). These results present evidence for spectral shape contrast for instrument sounds elicited by unrelated nonspeech contexts.

The relationship between responses collected in the questionnaire and the difference in 50% crossovers across filtered conditions was not significant ($r^2 = .43$, n.s.). Again, analyses of alternative nonlinear relationships between performance and musical experience also yielded no reliable fits to the data. Finally, no listener reported having previously heard the musical selection, nor could they identify its composer.

Although patterns of results are comparable, further analyses reveal the extent to which listeners' responses remained constant as a function of context source, which was the only difference between Experiments 1 (speech) and 2 (music). Similar performance would argue against both the importance of context and target being similar and the importance of listeners' familiarity with the materials. Crossover points from both experiments were entered into a mixed-design, 2 (experiment/context source: speech, music; between-subjects factor) × 3 (condition of context filtering: French horn − saxophone, saxophone − French horn, and unfiltered; within-subjects factor) ANOVA. The main effect of context source was not significant [$F(1,31) = 1.67$, n.s.], nor was the interaction between context filtering and experiment [$F(1.41,43.64) = 0.79$, n.s., with Greenhouse–Geiser sphericity correction]. Performance across experiments did not vary as a function of context source (i.e., music vs. speech), despite marked differences in acoustics and phenomenal experience.

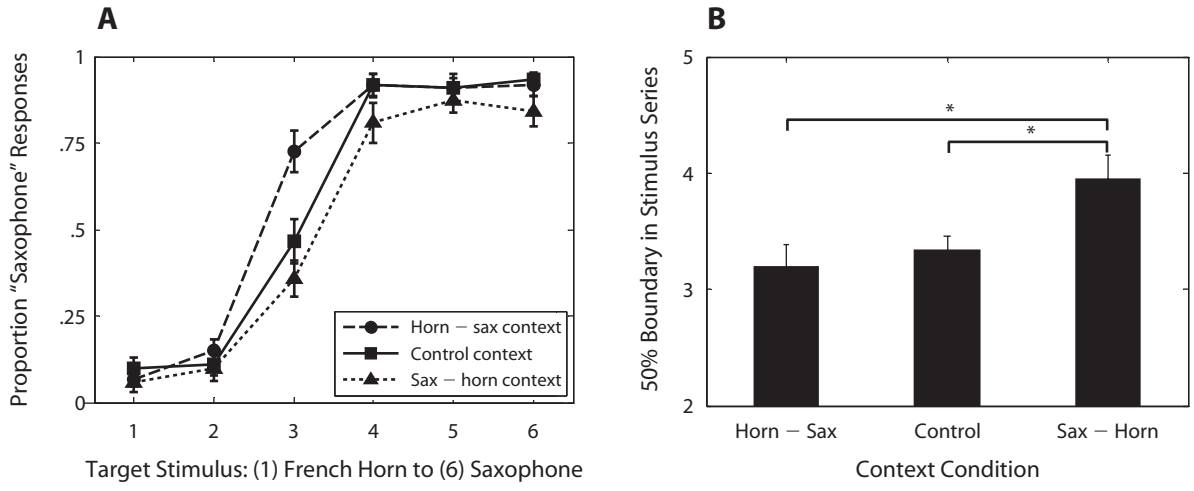

Figure 5. Results from 15 listeners in Experiment 2, using the same notation and symbols as used in Figure 4. (A) Proportion "saxophone" responses as a function of target stimulus. Steps along the abscissa correspond to the series of target sounds, with 1 corresponding to the French horn endpoint and 6 to the tenor saxophone endpoint. Standard error bars are depicted for each of the 18 trial types. (B) Mean 50% crossovers for each context condition, using the same formatting as Figure 4B. Error bars denote standard errors of the means. $^*p < .05$.

Discussion

Results demonstrated that perceptual calibration to reliable spectral properties generalizes to sounds that are not speech and are much less familiar to listeners than speech sounds are. Perception of instrument sounds was affected in a contrastive fashion relative to reliable spectral characteristics of a filtered musical passage. Listeners were more likely to perceive a saxophone when the preceding context was processed by the French horn − saxophone difference filter than when the preceding context was processed by the saxophone − French horn difference filter. Effects of preceding context were significant, despite no apparent source similarity between the Schubert String Quintet and target instrument spectra. This finding extends those by Watkins (1991) and Kiefte and Kluender (2008), in which both context and target not only were speech, but in some conditions, were spoken by the same person. Results also extend Holt's (2005, 2006a, 2006b) findings, in which spectrally sparse, statistically controlled distributions of sine-wave tones elicited contrast effects in listeners' identification of target stimuli as /da/ or /ga/. Here, reliable characteristics of the musical context (i.e., long-term spectra as a result of filtering), in which five string instruments played concurrently, elicited contrast effects in identification of target sounds from an entirely different class of instruments on physical, temporal, and spectral dimensions. In addition, perceptual adjustment for spectral characteristics of the listening context occurred even when listeners had relatively little experience with either context or target prior to the experiment.

GENERAL DISCUSSION

Consistent with the findings of Ladefoged and Broadbent (1957), Watkins (1991), Watkins and Makin (1994), Kiefte and Kluender (2008), and Alexander and Kluender (2009), the experiments reported here showed that reliable spectral characteristics of preceding context influence perception of subsequent target sounds. Those previous studies examined perception of highly familiar speech sounds with corresponding familiar spectral envelopes. By substituting musical and music-like stimuli for speech sounds, effects of reliable spectral characteristics were examined in sounds with which listeners have considerably less experience. Regardless of source, perception calibrated to predictable characteristics of listening context, consequently becoming more sensitive to changes in spectral composition of subsequent target items.

Calibration to reliable spectral characteristics was demonstrated for perception of modified musical instrument sounds following both music and speech. Despite no obvious relation between context and target, results demonstrate that perceptual consequences of reliable spectral characteristics occur for both unfamiliar context and unfamiliar target sounds. Processes responsible for these effects are sensitive primarily to reliable spectral characteristics and are apparently insensitive to other differences that may exist between contexts and targets, including differences in fundamental frequencies and in spectral compositions. Furthermore, the auditory system appears remarkably indifferent to sources or to changes between sources, regardless of whether they are highly familiar or relatively novel, when calibrating to both spectrally broad and local reliable spectral properties.

The present findings illuminate ways through which the auditory system calibrates for different listening conditions. Considered more broadly, these studies, in which listeners identify sounds in the context of other sounds, are more representative of everyday listening than are typical experiments concerning perception of isolated sounds under near-ideal conditions using headphones in soundproof booths. Outside the laboratory, spectra of sounds to which one is exposed are almost always colored by the environment. Energy at some frequencies is amplified by acoustic reflective properties of surfaces, whereas energy at other frequencies is attenuated by acoustically absorbent materials.

Patterns of performance in the current auditory experiments bear a striking similarity to the phenomenon of visual color constancy. In order to achieve color constancy, the visual system must somehow extract reliable properties of the spectrum of illumination across the entire image in order to determine inherent spectral properties of objects within the scene (e.g., Boynton, 1988; Churchland & Sejnowski, 1988; Foster et al., 1997). Furthermore, this calibration to the spectrum of illumination (e.g., Arend & Reeves, 1986) requires both light adaptation (Bramwell & Hurlbert, 1996; von Kries, 1905; Whittle, 1996) and contrast adaptation (Brown & MacLeod, 1997; Webster & Mollon, 1995).

Much as in the present experiments, the spectrum of illumination can be considered to be a filter that is reliably imposed on the full context of viewing, and perception of constant color is maintained by relative differences between the spectral composition of the object being viewed versus reliable spectral characteristics of the viewing context. As is the case with visual color constancy, calibration to listening context requires briefly becoming accustomed to the listening context.

There is one way in which calibration to listening context can be distinguished from visual color constancy. Kiefte and Kluender (2008) demonstrated that auditory perception calibrates to both gross spectral shape (tilt) and to relatively detailed spectral properties (spectral peak corresponding to F_2). By contrast, the visual system is not as successful when the illumination spectrum includes local spectral prominences or peaks. For example, fluorescent lights (which have multiple spectral peaks) and narrowband illumination (such as that from mercury or sodium vapor lights; Boynton and Purl, 1989; von Fieandt, Ahonen, Jarvinen, & Lian, 1964) compromise the ability of viewers to maintain color constancy. Unlike the visual system, the auditory system appears to be relatively more adept at compensating for local spectral perturbations.

It is sensible to assume that part of the difference between vision and audition results from different evolutionary pressures on the two systems, owing to optic and acoustic ecologies, and from differences in the structure of visual and auditory systems. With regard to optical ecology, it has been demonstrated that combinations of as few

as three basis functions can describe over 99% of spectral reflectances of Munsell color chips (Cohen, 1964; Maloney, 1986) and of natural objects (Maloney, 1986). The present authors are not aware of analogous analyses for acoustic ecology; however, if one examines only speech sounds as a subset of the total acoustic ecology (Maddieson, 1984), it is readily apparent that variance across speech sounds alone cannot be captured with such sparse descriptors. For example, at least 10 discrete cosine transform coefficients are required to classify simple spectrally constant vowel sounds in a fashion nearly comparable to human performance (Zahorian & Jagharghi, 1993). As for the comparative structures of visual and auditory systems, human color vision is accomplished using only four receptor classes (rods and cones), of which three (cones) are responsible for transducing most spectral information in daylight. By contrast, the human auditory system uses an array of about 7,000 transducers (inner hair cells; 3,500 per ear) to encode spectral composition. To the extent that effective hearing requires more detailed spectral analysis, it follows that the ability to effectively encode predictable characteristics of the acoustic environment, both spectrally gross and local, is essential to increasing sensitivity of the auditory system to informative and ecologically significant spectral differences.

Physiological mechanisms through which the auditory system calibrates for characteristics of acoustic context have not been extensively investigated and are not yet understood. Primary auditory cortex (AI) neurons decode spectral shape with respect to both broad and narrow complex spectral shapes (Barbour & Wang, 2003). Furthermore, AI neurons are sensitive to the relative probabilities of pure tones of different frequencies in an extended sequence of tones (Ulanovsky, Las, & Nelken, 2003). Even in the inferior colliculus of the brainstem, single neurons have increased excitability in response to changes from predictable acoustic patterns (Dean, Harper, & McAlpine, 2005; Dean, Robinson, Harper, & McAlpine, 2008; Malmierca, Cristaudo, Pérez-González, & Covey, 2009; Pérez-González, Malmierca, & Covey, 2005).

Yet it has been hypothesized that, lower in the auditory system, projections from superior olive to outer hair cells (medial olivary complex) provide adjustments of basilar membrane tuning to improve resolution of signals against background noise (Kirk & Smith, 2003); one could speculate as to whether these projections could be sufficiently sophisticated to account for effects of both spectrally broad and local contexts.

Outside the laboratory, the present findings are significant for listening in typical conditions in which sounds arriving at the ear are structured by properties of the environment. Of particular interest may be concert hall acoustics and perception of orchestral and other types of music. Listeners in Experiment 2 heard a short musical selection followed by a target instrument stimulus whose identification clearly was affected by reliable, long-term spectral characteristics of the preceding context. This finding offers insight into perception of orchestral music, where musical instruments from string, brass, and woodwind sections play in concert and in succession, and resonance characteristics of concert halls provide consistent spectral filtering. In related studies, experience with room acoustics was shown to result in perceptual calibration to regularities in reverberations modulating perceptual judgments (e.g., Freyman, Clifton, & Litovsky, 1991; Watkins, 2005a, 2005b).

Across the studies presented here, data illustrate how the auditory system calibrates to reliable properties of a listening context in ways that enhance sensitivity to change. These findings are consistent with very general principles concerning how perceptual systems work. Perceptual systems respond predominantly to change; they do not record absolute levels—whether of loudness, pitch, brightness, or color—and this has been demonstrated perceptually in every sensory domain (e.g., Kluender, Coady, & Kiefte, 2003). This sensitivity to change not only increases the effective dynamic range of biological systems, it also increases the amount of information conveyed between environment and organism (Kluender & Alexander, 2008; Kluender & Kiefte, 2006).

AUTHOR NOTE

This work was supported by a grant from the Social Sciences and Humanities Research Council to M.K. and by Grant DC 004072 from the National Institutes of Deafness and Communication Disorders to K.R.K. The authors are grateful to Lynne Nygaard and three anonymous reviewers for very helpful suggestions in response to an earlier version of this article, and we extend thanks to Stephanie Jacobs, Kathrine Allie, and Kyira Hauer for assistance in participant recruitment and testing. Address correspondence to C. E. Stilp, Department of Psychology, University of Wisconsin, 1202 West Johnson St., Madison, WI 53706 (e-mail: cestilp@wisc.edu).

REFERENCES

ALEXANDER, J. M., & KLUENDER, K. R. (2009). *Temporal properties of auditory perceptual calibration*. Manuscript submitted for publication.

ARAVAMUDHAN, R., LOTTO, A. J., & HAWKS, J. W. (2008). Perceptual context effects of speech and nonspeech sounds: The role of auditory categories. *Journal of the Acoustical Society of America*, **124**, 1695-1703. doi:10.1121/1.2956482

AREND, L., & REEVES, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A*, **3**, 1743-1751. doi:10.1364/JOSAA.3.001743

BARBOUR, D. L., & WANG, X. (2003). Contrast tuning in auditory cortex. *Science*, **299**, 1073-1075. doi:10.1126/science.1080425

BOERSMA, P., & WEENINK, D. (2007). Praat: Doing phonetics by computer (Version 4.5.12) [Computer software]. Retrieved January 31, 2007, from www.praat.org.

BOYNTON, R. M. (1988). Color vision. *Annual Review of Psychology*, **39**, 69-100.

BOYNTON, R. M., & PURL, K. F. (1989). Categorical colour perception under low-pressure sodium lighting with small amounts of added incandescent illumination. *Lighting Research & Technology*, **21**, 23-27. doi:10.1177/096032718902100104

BRAMWELL, D. I., & HURLBERT, A. C. (1996). Measurements of colour constancy by using forced-choice matching technique. *Perception*, **25**, 229-241. doi:10.1068/p250229

BROWN, R. O., & MACLEOD, D. I. A. (1997). Color appearance depends on the variance of surround colors. *Current Biology*, **7**, 844-849. doi:10.1016/S0960-9822(06)00372-1

CHURCHLAND, P. S., & SEJNOWSKI, T. J. (1988). Perspectives on cognitive science. *Science*, **242**, 741-745. doi:10.1126/science.3055294

COHEN, J. (1964). Dependency of the spectral reflectance curves of the Munsell color chips. *Psychonomic Science*, **1**, 369-370.

DEAN, I., HARPER, N. S., & MCALPINE, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience*, **8**, 1684-1689. doi:10.1038/nn1541

DEAN, I., ROBINSON, B. L., HARPER, N. S., & MCALPINE, D. (2008). Rapid neural adaptation to sound level statistics. *Journal of Neuroscience*, **28**, 6430-6438. doi:10.1523/JNEUROSCI.0470-08.2008

FOSTER, D. H., AMANO, K., & NASCIMENTO, S. M. C. (2001). Colour constancy from temporal cues: Better matches with less variability under fast illuminant changes. *Vision Research*, **41**, 285-293. doi:10.1016/S0042-6989(00)00239-X

FOSTER, D. H., NASCIMENTO, S. M. C., CRAVEN, B. J., LINNELL, K. J., CORNELISSEN, F. W., & BRENNER, E. (1997). Four issues concerning colour constancy and relational colour constancy. *Vision Research*, **37**, 1341-1345. doi:10.1016/S0042-6989(96)00285-4

FREYMAN, R. L., CLIFTON, R. K., & LITOVSKY, R. Y. (1991). Dynamic processes in the precedence effect. *Journal of the Acoustical Society of America*, **90**, 874-884. doi:10.1121/1.401955

GREY, J. M. (1975). *An exploration of musical timbre*. Unpublished doctoral dissertation, Stanford University.

HOLT, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, **16**, 305-312. doi:10.1111/j.0956-7976.2005.01532.x

HOLT, L. L. (2006a). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *Journal of the Acoustical Society of America*, **120**, 2801-2817. doi:10.1121/1.2354071

HOLT, L. L. (2006b). Speech categorization in context: Joint effects of nonspeech and speech precursors. *Journal of the Acoustical Society of America*, **119**, 4016-4026. doi:10.1121/1.2195119

KIEFTE, M., & KLUENDER, K. R. (2008). Absorption of reliable spectral characteristics in auditory perception. *Journal of the Acoustical Society of America*, **123**, 366-376. doi:10.1121/1.2804951

KIRK, E. C., & SMITH, D. W. (2003). Protection from acoustic trauma is not a primary function of the medial olivocochlear efferent system. *Journal of the Association for Research in Otolaryngology*, **4**, 445-465. doi:10.1007/s10162-002-3013-y

KLUENDER, K. R., & ALEXANDER, J. M. (2008). Perception of speech sounds. In A. I. Basbaum, A. Kaneko, G. M. Shepard, & G. Westheimer (Series Eds.) & P. Dallos & D. Ortel (Vol. Eds.), *The senses: A comprehensive reference. Vol. 3: Audition* (pp. 829-860). San Diego: Academic Press.

KLUENDER, K. R., COADY, J. A., & KIEFTE, M. (2003). Sensitivity to change in perception of speech. *Speech Communication*, **41**, 59-69. doi:10.1016/S0167- 6393(02)00093-6

KLUENDER, K. R., & KIEFTE, M. (2006). Speech perception within a biologically-realistic information-theoretic framework. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 153-199). Amsterdam: Elsevier.

KRUMHANSL, C. (1989). Why is musical timbre so hard to understand? In S. Nielzén & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43-53). Amsterdam: Excerpta Medica.

LADEFOGED, P., & BROADBENT, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, **29**, 98-104. doi:10.1121/1.1908694

LAND, E. H. (1983). Recent advances in retinex theory and some implications for cortical computations: Color vision and the natural image. *Proceedings of the National Academy of Sciences*, **80**, 5163-5169.

MADDIESON, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

MALMIERCA, M. S., CRISTAUDO, S., PÉREZ-GONZÁLEZ, D., & COVEY, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *Journal of Neuroscience*, **29**, 5483-5493. doi:10.1523/JNEUROSCI.4153-08.2009

MALONEY, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America A*, **3**, 1673-1683. doi:10.1364/JOSAA.3.001673

MCADAMS, S., WINSBERG, S., DONNADIEU, S., DE SOETE, G., & KRIMPHOFF, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, **58**, 177-192. doi:10.1007/BF00419633

MCCANN, J. J., MCKEE, S. P., & TAYLOR, T. H. (1976). Quantitative studies in retinex theory: A comparison between theoretical predictions and observer responses to the "color Mondrian" experiments. *Vision Research*, **16**, 445-458. doi:10.1016/0042-6989(76)90020-1

MCCULLAGH, P., & NELDER, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

MILLER, G. A., & NICELY, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, **27**, 338-352. doi:10.1121/1.1907526

NASSAU, K. (1983). *The physics and chemistry of color: The fifteen causes of color.* New York: Wiley.

OPOLKO, F., & WAPNICK, J. (1989). *McGill University master samples user's manual*. Montreal: McGill University, Faculty of Music.

PÉREZ-GONZÁLEZ, D., MALMIERCA, M. S., & COVEY, E. (2005). Novelty detector neurons in the mammalian auditory midbrain. *European Journal of Neuroscience*, **22**, 2879-2885. doi:10.1111/j.1460-9568.2005.04472.x

PISONI, D. B. (1987). General discussion of session 3: Dynamic aspects. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception, Series D* (pp. 264-267). Dordrecht: Nijhoff.

STEPHENS, J. D. W., & HOLT, L. L. (2003). Preceding phonetic context affects perception of nonspeech. *Journal of the Acoustical Society of America*, **114**, 3036-3039. doi:10.1121/1.1627837

ULANOVSKY, N., LAS, L., & NELKEN, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, **6**, 391-398. doi:10.1121/1.1627837

VON FIEANDT, K., AHONEN, L., JARVINEN, J., & LIAN, A. (1964). Color experiments with modern sources of illumination. *Annales Academiae Scientiarum Fennicae: Series B*, **134**, 3-89.

VON KRIES, J. (1905). Die Gesichtsempfindungen. In W. Nagel, *Physiologie der Sinne: Vol. 3. Handbuch der Physiologie des Menschen* (pp. 109-282). Braunschweig: Vieweg & Sohn.

WATKINS, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America*, **90**, 2942-2955. doi:10.1121/1.401769

WATKINS, A. J. (2005a). Perceptual compensation for effects of echo and of reverberation on speech identification. *Acta Acustica*, **91**, 892-901.

WATKINS, A. J. (2005b). Perceptual compensation for effects of reverberation in speech identification. *Journal of the Acoustical Society of America*, **118**, 249-262. doi:10.1121/1.1923369

WATKINS, A. J., & MAKIN, S. J. (1994). Perceptual compensation for speaker differences and for spectral-envelope distortion. *Journal of the Acoustical Society of America*, **96**, 1263-1282. doi:10.1121/1.410275

WEBSTER, M. A., & MOLLON, J. D. (1995). Colour constancy influenced by contrast adaptation. *Nature*, **373**, 694-698. doi:10.1038/373694a0

WHITTLE, P. (1996). Perfect von Kries contrast colours. *Perception*, **25**(Suppl.), 16.

ZAHORIAN, S. A., & JAGHARGHI, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *Journal of the Acoustical Society of America*, **94**, 1966-1982. doi:10.1121/1.407520

ZAIDI, Q., SPEHAR, B., & DEBONET, J. (1997). Color constancy in variegated scenes: Role of low-level mechanisms in discounting illumination changes. *Journal of the Optical Society of America A*, **14**, 2608-2621. doi:10.1364/JOSAA.14.002608

NOTES

1. Aravamudhan, Lotto, and Hawks (2008) reported obtaining evidence for speech context affecting identification of nonspeech target sounds, but results are limited by two shortcomings. Nonspeech targets were actually sine-wave analogues of speech sounds, and the context effect was observed only following thorough training with feedback on the target stimuli. These shortcomings obscure the extent to which this evidence supports the generality of contrast effects.

2. Alexander and Kluender (2009) demonstrated that effects of preceding context asymptote with 1-sec and longer duration. Both Experiments 1 and 2 also were conducted with 2.5-sec contexts, and results were the same as reported here.

3. To address the possibility that observed contrast effects could be attributed to specific spectral properties at the offset of sentence context, an additional experiment was conducted utilizing time-reversed versions of the contexts presented in Experiment 1. Data from 25 listeners who did not participate in any other experiments produced results comparable to those reported for Experiment 1.

(Manuscript received April 17, 2009;
revision accepted for publication September 4, 2009.)