



# Natural music context biases musical instrument categorization

Joshua M. Lanning<sup>1</sup> · Christian Stilp<sup>1</sup>

Published online: 19 February 2020  
© The Psychonomic Society, Inc. 2020

## Abstract

Perception of sounds occurs in the context of surrounding sounds. When spectral properties differ between earlier (context) and later (target) sounds, categorization of later sounds becomes biased through spectral contrast effects (SCEs). Past research has shown SCEs to bias categorization of speech and music alike. Recent studies have extended SCEs to naturalistic listening conditions when the inherent spectral composition of (unfiltered) sentences biased speech categorization. Here, we tested whether natural (unfiltered) music would similarly bias categorization of French horn and tenor saxophone targets. Preceding contexts were either solo performances of the French horn or tenor saxophone (unfiltered; 1 second duration in Experiment 1, or 3 seconds duration in Experiment 2) or a string quintet processed to emphasize frequencies in the horn or saxophone (filtered; 1 second duration). Both approaches produced SCEs, producing more “saxophone” responses following horn / horn-like contexts and vice versa. One-second filtered contexts produced SCEs as in previous studies, but 1-second unfiltered contexts did not. Three-second unfiltered contexts biased perception, but to a lesser degree than filtered contexts did. These results extend SCEs in musical instrument categorization to everyday listening conditions.

**Keywords** Music perception · Context effects · Perceptual categorization · Hearing · Spectral contrast effects

## Introduction

All perception takes place in context. Objects and events in the environment are not perceived in isolation, but relative to other objects and events as well as recent perceptual experience. When surrounding (context) stimuli differ from a given (target) stimulus, perceptual systems magnify this difference, resulting in a contrast effect. Contrast effects occur in all sensory modalities (von Békésy, 1967; Warren, 1985) and have a very broad influence, affecting perception of simple stimuli such as tones (Christman, 1954) and lines (Gibson, 1937) to more complex stimuli such as faces (Webster & MacLeod,

2011), emotions (Bestmeyer, Rouger, DeBruine, & Belin, 2010), and gender (Troje, Sadr, Geyer, & Nakayama, 2006).

Contrast effects play a significant role in speech perception (for reviews, see Kluender, Coady, & Kieft, 2003; Stilp, 2019b). When spectral properties differ between earlier (context) and later (target) sounds, this can produce spectral contrast effects (SCEs) that bias categorization of later sounds. For example, when context sounds are spectrally similar to /o/ (higher frequency first formant), listeners are more likely to categorize the subsequent target vowel as /u/ (lower frequency first formant) and vice versa. SCEs have been repeatedly shown to influence speech perception (e.g., Holt, 2006; Ladefoged & Broadbent, 1957; Sjerps, Zhang, & Peng, 2018; Stilp, 2019b; Stilp, Anderson, & Winn, 2015; Stilp & Assgari, 2019; Watkins, 1991). This influence extends to perception of other complex sounds such as musical instruments (Stilp, Alexander, Kieft, & Kluender, 2010). This parallel between speech and music perception deepens as SCE magnitudes vary in orderly ways for perception of vowels (Stilp et al., 2015), consonants (Stilp & Assgari, 2017), and musical instruments (Frazier, Assgari, & Stilp, 2019).

Historically, investigations of SCEs have used carefully acoustically controlled stimuli, often testing multiple filtered renditions of a single context stimulus. While providing great experimental control, questions remain regarding how SCEs

---

**Significance statement** Perception of a sound is always in the context of other sounds. This is especially the case when perception accentuates the difference between a sound’s frequency composition and that of earlier sounds. These “spectral contrast effects” have been reported in speech and music perception alike. Here, these effects are produced by musical excerpts that already contained the relevant frequency information without any additional filtering or editing. This extends these context effects to more naturalistic conditions such as everyday music listening.

---

✉ Christian Stilp  
christian.stilp@louisville.edu

<sup>1</sup> Department of Psychological and Brain Sciences, University of Louisville, 317 Life Sciences Building, Louisville, KY 40292, USA

contribute to perception in more variable (and naturalistic) listening conditions. Recently, Stilp and Assgari (2019) observed SCEs in vowel categorization following highly controlled filtered sentences (as in previous studies) as well as sentences that naturally possessed the desired spectral properties without filtering, significantly enhancing the ecological validity of SCEs. Additionally, SCEs produced by unfiltered sentences were smaller than SCEs produced by filtered sentences, shedding considerable light on the precise degree to which these effects shape everyday perception. While SCEs have patterned similarly in categorization of speech and musical instruments thus far, it is unknown whether SCEs in music perception similarly extend to more naturalistic listening conditions. The present study tested whether unfiltered contexts (i.e., musical segments that already possess the desired spectral properties) produce SCEs in the perception of music in the same manner as was previously reported for perception of speech. On each trial, listeners heard a musical context (French horn or tenor saxophone excerpt, or a string quintet filtered to emphasize frequencies of the horn or saxophone) before categorizing an instrument from a French horn–tenor saxophone target continuum.

## Method

### Participants

Forty-seven undergraduate students from the University of Louisville participated in the experiments ( $n = 25$  in Experiment 1,  $n = 22$  in Experiment 2, with no one completing both). Each participant self-reported normal hearing and was compensated with course credit.

### Stimuli

#### Targets

Target stimuli were the same as those used in Stilp et al. (2010) and Frazier et al. (2019). Recordings of the French horn and tenor saxophone playing the note G3 (196 Hz) were taken from the McGill University Musical Samples database (Opolko & Wapnick, 1989). Three consecutive pitch pulses (15.31 ms) of constant amplitude were excised at zero-crossings from each recording, iterated to 140 ms total duration in Praat (Boersma & Weenink, 2019), and processed by 5-ms linear onset and offset ramps. Stimuli were then proportionately mixed along a six-step continuum so that the amplitude of one instrument was +30, +18, +6, −6, −18, or −30 dB relative to the other. Finally, waveforms were low-pass filtered at a 10-kHz cutoff. Instrument mixing and filtering was performed in MATLAB.

### Contexts

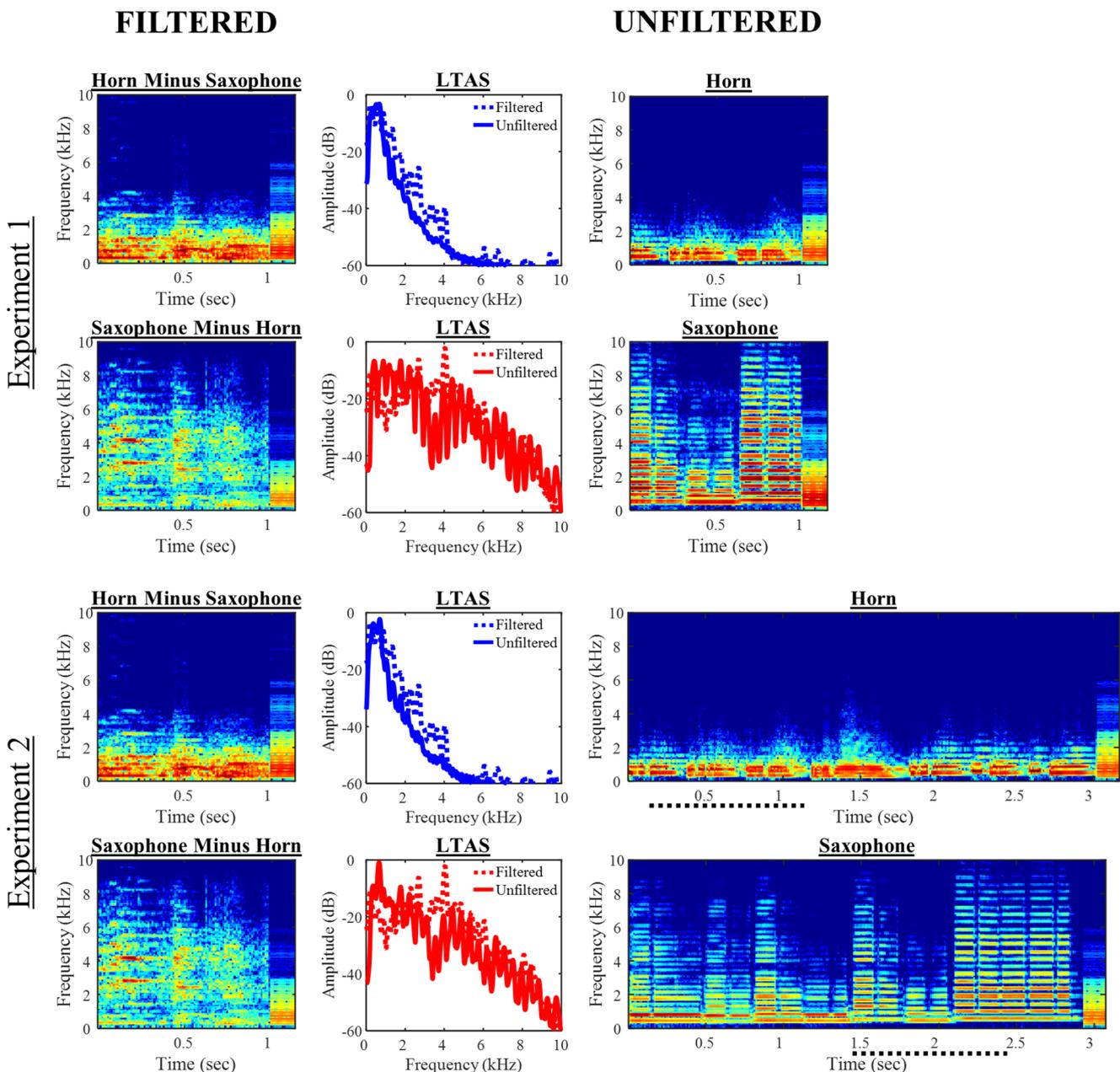
**Unfiltered** Context stimuli were two musical instrument solos drawn from YouTube videos. The saxophone context was a 2,938-ms excerpt of Ivan Renta playing tenor saxophone for “Profiles in Greatness.” The horn context was a 3,032-ms excerpt of Sarah Willis of the Berlin Philharmonic playing French horn in the seldom-used unstopped playing style (i.e., hand not placed in the horn’s bell to alter its spectrum, in order to match the unstopped horn target stimuli). Different musical selections were selected in order to model the high acoustic variability of everyday perception, and to parallel Stilp and Assgari (2019), who presented two different sentences in their unfiltered conditions. One-second excerpts of these stimuli were presented as contexts in Experiment 1, and were presented in their entireties in Experiment 2 (see Fig. 1).

**Filtered** The context stimulus was a 1-second excerpt from Franz Schubert’s string quintet in C major, “Allegretto,” taken from a compact disc. As in Stilp et al. (2010) and Frazier et al. (2019), endpoint French horn and tenor saxophone stimuli were analyzed to create spectral envelope difference (SED) filters (Watkins, 1991). Spectral envelopes for each instrument were derived from 512-point Fourier transforms, and smoothed using a 256-point Hamming window with 128-point overlap. Spectral envelopes were equated for peak power, then subtracted from one another in both directions (horn minus saxophone, saxophone minus horn). A finite impulse response was obtained for each SED using inverse Fourier transform. The string quintet context was processed by each SED filter (see Fig. 1).

### Procedure

Both experiments followed the same procedure. After obtaining informed consent, each participant entered a sound-attenuating booth (Acoustic Systems, Inc., Austin, TX). All sounds were D/A converted by RME HDSPe AIO sound cards (Audio AG, Haimhausen, Germany) on a personal computer and passed through a programmable attenuator (TDT PA4, Tucker-Davis Technologies, Alachua, FL) and headphone buffer (TDT HB6) before being presented diotically at 70 dB sound pressure level over circumaural headphones (Beyerdynamic DT-150, Beyerdynamic Inc. USA, Farmingdale, NY).

A custom MATLAB script guided participants through the experiment, which consisted of four phases. The first phase was exposure, in which each instrument endpoint (French horn or tenor saxophone) was played twice along with its verbal label. The second phase was practice, where participants completed 100 trials identifying endpoints from the French horn–tenor saxophone series. Participants were required to achieve at least 90% correct, which all achieved.



**Fig. 1** Stimuli presented in Experiment 1 (top two rows) and Experiment 2 (bottom two rows). Spectrograms of sample trials with filtered contexts (string quintet processed by horn-minus-saxophone filter or saxophone-minus-horn filter) are in the left column. Spectrograms of sample trials with unfiltered contexts (horn, saxophone) are in the right column. Dotted lines underneath unfiltered spectrograms in Experiment 2 indicate 1-second segments that were excised and presented as contexts in

Experiment 1. The long-term average spectra (LTAS) of filtered (dashed lines) and unfiltered (solid lines) contexts depicted in each row are overlaid in the center column for comparison. All spectrograms depict the same instrument target (step number 3 out of 6); sample trials are posted online at <https://louisville.edu/psychology/stilp/lab/demos>. (Color figure online)

The third phase was the main experiment. On each trial, participants heard a context stimulus followed by a musical instrument target. Participants clicked the mouse to indicate whether the target sounded more like a French horn or a tenor saxophone. Each experiment consisted of two blocks (filtered contexts, unfiltered contexts), with each block presenting 120 trials [2 contexts (frequencies favoring the horn or the saxophone) × 6 target instruments × 10 repetitions]. Stimuli were

randomized within each block, and blocks were tested in counterbalanced orders across participants. The final phase was a five-question survey of listeners’ self-rated musical performance ability, years of musical training and performance experience, and other relevant experience (same questionnaire as in Stilp et al., 2010, and Frazier et al., 2019). The entire session took approximately 30 minutes to complete.

## Results: Experiment 1

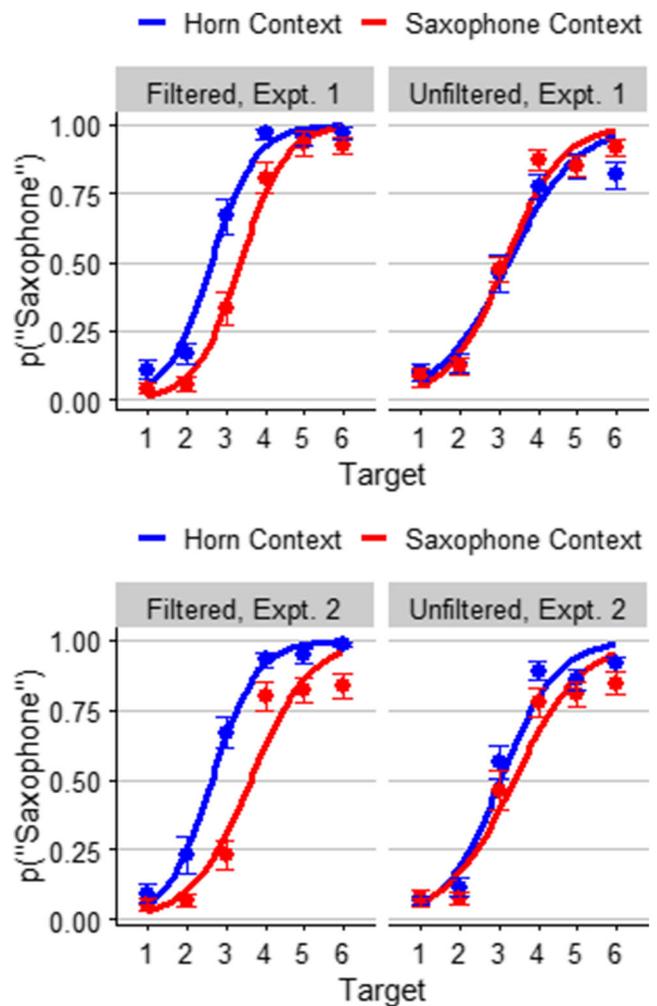
While all participants successfully completed the practice phase, they were required to maintain 90% accuracy on instrument endpoints throughout the main experiment in order to have their data included in analyses. Surprisingly few participants met this criterion (9/25 in Experiment 1; 7/22 in Experiment 2), so a laxer criterion of 80% accuracy was adopted (12/25 met this criterion in Experiment 1; 13/22 in Experiment 2). While smaller, a sample size of 12 listeners retained 83% statistical power to detect SCEs [Cohen's  $d = 0.80$  modeled after SCEs in previous filtered conditions; (Frazier et al., 2019; Stilp et al., 2010);  $\alpha = .05$ , one-tailed dependent-samples  $t$  test; as calculated using G\*Power; Faul, Erdfelder, Buchner, & Lang, 2009].

For each experiment, responses were analyzed with a mixed-effects logistic model in R (R Development Core Team, 2016) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). The dependent variable was binary ("horn" or "saxophone" responses coded as 0 and 1, respectively). Fixed effects in the model included Target (coded as a continuous variable from 1 to 6 then mean-centered), Context (sum coded with two levels: horn/horn-minus-saxophone as  $-0.5$  and saxophone/saxophone-minus-horn as  $+0.5$ ), Condition (sum coded with two levels: filtered as  $-0.5$  and unfiltered as  $+0.5$ ), and all interactions between fixed effects. Random slopes were included for each fixed main effect, and a random intercept of participant was included.

Results from Experiment 1 are presented in Table 1 and the top row of Fig. 2. Listeners responded "saxophone" more often overall (significant intercept), with each rightward step along the target instrument continuum (main effect of Target), and following filtered contexts (negative main effect of Condition). Listeners responded "saxophone" less often when context frequencies were changed from predominantly lower frequencies (favoring the horn) to higher frequencies (favoring the saxophone; negative main effect of Context), consistent with the predicted direction of SCEs. SCE magnitudes significantly differed across unfiltered and filtered conditions (interaction between Context and Condition). SCEs were then

**Table 1** Mixed-effects model results for Experiment 1. Levels sum-coded at  $-0.5$  are listed in parentheses.  $SE$  = standard error of the mean

|  | Estimate | $SE$  | $Z$      | $p$    |
|--|----------|-------|----------|--------|
| Intercept                                  | 0.645    | 0.158 | 4.079    | <.0001 |
| Target                                     | 1.503    | 0.170 | 8.846    | <.0001 |
| Context (horn/horn-minus-sax)              | $-0.567$ | 0.180 | $-3.147$ | .002   |
| Condition (filtered)                       | $-0.485$ | 0.178 | $-2.717$ | .007   |
| Target $\times$ Context                    | 0.113    | 0.093 | 1.215    | .224   |
| Target $\times$ Condition                  | $-0.500$ | 0.094 | $-5.338$ | <.0001 |
| Context $\times$ Condition                 | 1.484    | 0.234 | 6.352    | <.0001 |
| Target $\times$ Context $\times$ Condition | 0.259    | 0.184 | 1.405    | .160   |



**Fig. 2** Mixed-effects model fits to behavioral results from Experiments 1 (top row) and 2 (bottom row). The probability of responding "saxophone" is plotted as a function of the target instrument (1 = French horn endpoint, 6 = tenor saxophone endpoint). Filtered conditions appear at left (blue = horn-minus-saxophone-filtered context, red = saxophone-minus-horn-filtered context), unfiltered conditions appear at right (blue = horn context, red = saxophone context). Circles depict mean probabilities of a "saxophone" response to each target stimulus; error bars depict standard error of the mean. (Color figure online)

calculated for each listener in each condition as the percentage shift in "saxophone" responses following different contexts. SCEs were markedly larger following filtered contexts (mean SCE = 12.22% shift) than unfiltered contexts (mean =  $-3.19\%$  shift), which did not significantly differ from zero (one-tailed  $t$  test):  $t(11) = 1.29$ ,  $p = .11$ .

## Results: Experiment 2

Results from Experiment 2 are presented in Table 2 and the bottom row of Fig. 2. Listeners again responded "saxophone" more often as in Experiment 1 (significant intercept, main effects of Target and Condition). SCEs occurred again in

**Table 2** Mixed-effects model results for Experiment 2. Levels contrast-coded at  $-0.5$  are listed in parentheses. *SE* = standard error of the mean

|  | Estimate | <i>SE</i> | <i>Z</i> | <i>p</i> |
|--|----------|-----------|----------|----------|
| Intercept                                  | 0.533    | 0.168     | 3.175    | .002     |
| Target                                     | 1.399    | 0.151     | 9.286    | <.0001   |
| Context (horn/horn-minus-sax)              | -1.087   | 0.152     | -7.174   | <.0001   |
| Condition (filtered)                       | -0.266   | 0.132     | -2.015   | .044     |
| Target $\times$ Context                    | -0.323   | 0.087     | -3.733   | <.001    |
| Target $\times$ Condition                  | -0.244   | 0.086     | -2.834   | .005     |
| Context $\times$ Condition                 | 1.143    | 0.223     | 5.122    | <.0001   |
| Target $\times$ Context $\times$ Condition | 0.154    | 0.168     | 0.919    | .358     |

Experiment 2 (negative main effect of Context), and the significant Context  $\times$  Condition interaction indicates that effects were again larger following filtered (mean across participants = 17.31% shift in “saxophone” responses) than unfiltered contexts (mean = 6.15% shift). SCEs following unfiltered contexts in Experiment 2 were significantly larger than zero (one-sample *t* test),  $t(12) = 2.88$ ,  $p < .025$ , and larger than SCEs following unfiltered contexts in Experiment 1 (independent-samples *t* test),  $t(23) = 2.87$ ,  $p < .01$ .

No survey responses were significant predictors of SCE magnitudes in either experiment. This null result (also observed in Stilp et al., 2010, and Frazier et al., 2019) is possibly due to investigating musical experience in a sample of nonmusicians (mean years of performance experience  $\approx 2$  in Experiment 1,  $\approx 1$  in Experiment 2).

## Discussion

Contrast effects are fundamental to perception in all sensory modalities (von Békésy, 1967; Warren, 1985). The challenge of rampant stimulus variability is lessened by emphasizing *differences* between stimuli, which highlights new information that can inform adaptive behavior (Kluender & Alexander, 2007). Perception of speech and nonspeech sounds is widely influenced by spectral contrast effects (SCEs; e.g., Christman, 1954; Kingston et al., 2014; Ladefoged & Broadbent, 1957; Stilp et al., 2010; Stilp et al., 2015; for reviews, see Kluender et al., 2003; Stilp, 2019b). These demonstrations generally used highly acoustically controlled stimuli, which affords good experimental control, but does not necessarily model the acoustic variability present in everyday perception. Naturalistic (unfiltered) sentence contexts were recently shown to influence vowel categorization via SCEs, but to a lesser degree than filtered renditions of a single sentence context (Stilp & Assgari, 2019). The present experiments extended this paradigm to musical instrument categorization, which thus far has exhibited strong parallels to speech categorization in terms of SCEs (Frazier et al., 2019; Stilp et al., 2010).

In Experiment 1, unfiltered musical contexts failed to produce an SCE (see Fig 2, top right). Simple acoustic descriptions fail to explain this result, as unfiltered and filtered contexts were matched in long-term average spectra (see Fig 1) and duration (1 second). This null result does not appear to be stimulus specific, as it was replicated with new participants using different 1-second excerpts from the same 3-second musical contexts (results not shown, but data are available online; see Open Practices Statement). In Experiment 2, 3-second unfiltered contexts did bias instrument categorization via an SCE (see Fig. 2, bottom right). Long-term average spectra were again well-matched to those of filtered contexts (and the contexts tested in Experiment 1; see Fig. 1), so this cannot explain different patterns of results across experiments. Instead, 1-second unfiltered musical contexts appear to be ineffective in biasing subsequent instrument categorization. Longer-duration contexts have been shown to produce larger spectral context effects (SCEs: Holt, 2006; auditory enhancement effects [which are related to SCEs; Stilp, 2019a]: Viemeister, 1980). The present experiments successfully extended SCEs in music perception to more naturalistic listening conditions (i.e., using unfiltered context stimuli), but also suggest that adequate sampling of the context is necessary for effects to manifest.

Critically, the present experiments compared how highly controlled filtered contexts and less-controlled unfiltered contexts elicited SCEs in musical instrument categorization. Filtered contexts produced SCEs in both experiments, replicating previous studies (Frazier et al., 2019; Stilp et al., 2010), but now served as the baseline for interpreting results from unfiltered conditions. Despite being matched to filtered contexts in duration (Experiment 1) and long-term average spectra (Experiments 1 and 2), unfiltered contexts produced smaller (if any) SCEs than filtered contexts in each experiment. This finding parallels Stilp and Assgari (2019), where unfiltered context sentences produced smaller and more variable SCEs in vowel categorization than filtered renditions of a single context sentence. In that study, each unfiltered block presented two different sentences sometimes spoken by two different talkers, similar to two different musical passages played by two different musical instruments (and musicians) presented here. Acoustic variability across the unfiltered contexts likely contributed to them producing smaller SCEs than filtered contexts, which directly challenges Holt’s (2006) claim that context variability does not affect SCEs. Holt’s contexts were sequences of pure tones with increasing variance in their frequencies, a rather restricted testing case considering the immense spectrotemporal variability of complex contexts such as music and speech.

Participants categorized musical instrument endpoints accurately ( $\geq 90\%$ ) in the practice session, but most struggled to maintain this performance in the main experiment, leading to a more lenient performance criterion ( $\geq 80\%$  correct). Some of

this difficulty might have arisen from a degree of stimulus uncertainty. Despite being instructed to categorize some horns and saxophones (targets) but not others (contexts), some participants categorized the context instrument rather than the target instrument. Also, endpoint stimuli might be more difficult to categorize accurately in context (following a musical passage) than in isolation, but participants in Frazier et al. (2019) were largely successful at both (15/17 participants met the 90% criterion). Further research is needed to identify the reason(s) for this difficulty.

In conclusion, the present experiments tested the influence of filtered or unfiltered musical contexts on musical instrument categorization. Following Stilp and Assgari (2019), this experimental paradigm deliberately sacrificed considerable acoustic control in order to test naturalistic (unfiltered) contexts that were more acoustically variable (as is commonly the case in everyday perception). Unfiltered musical passages biased perception of musical instruments (Experiment 2), but to a lesser degree than filtered contexts. These results parallel Stilp and Assgari (2019), who reported that unfiltered sentences biased vowel categorization via SCEs but to lesser degrees than filtered sentences did. Together, these results extend SCEs in speech and nonspeech perception to more naturalistic listening conditions while also better informing the precise degree to which SCEs influence everyday auditory perception.

**Acknowledgements** The authors wish to thank the Action Editor Meghan Clayards and two anonymous reviewers for helpful comments and suggestions. We also thank Scott Barrett, Ella Beilman, Rebecca Davis, Emily Dickey, Pratistha Thapa, and Sara Wardrip for their assistance with data collection. This study was presented as the first author's Culminating Undergraduate Experience in the Department of Psychological and Brain Sciences at the University of Louisville.

**Open practices statement** All data and analysis scripts are available in the Open Science Framework (<https://osf.io/dgbz9>).

## References

- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (R Package Version 1.1-7) [Computer software]. Retrieved from <http://cran.r-project.org/package=lme4>
- Bestelmeyer, P. E. G., Rouger, J., DeBruine, L. M., & Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition*, *117*(2), 217–223.
- Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer [Computer program]. Retrieved from <http://www.fon.hum.uva.nl/praat/>
- Christman, R. J. (1954). Shifts in pitch as a function of prolonged stimulation with pure tones. *The American Journal of Psychology*, *67*(3), 484–491.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.
- Frazier, J. M., Assgari, A. A., & Stilp, C. E. (2019). Musical instrument categorization is highly sensitive to spectral properties of earlier sounds. *Attention, Perception, & Psychophysics*, *81*(4), 1119–1126.
- Gibson, J. J. (1937). Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *Journal of Experimental Psychology*, *20*(6), 553–569.
- Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *Journal of the Acoustical Society of America*, *120*(5), 2801–2817.
- Kingston, J., Kawahara, S., Chambless, D., Key, M., Mash, D., & Watsky, S. (2014). Context effects as auditory contrast. *Attention, Perception, & Psychophysics*, *76*, 1437–1464.
- Kluender, K. R., & Alexander, J. M. (2007). Perception of speech sounds. In P. Dallos & D. Oertel (Eds.), *The senses: A comprehensive reference* (pp. 829–860). San Diego, CA: Academic Press.
- Kluender, K. R., Coady, J. A., & Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Communication*, *41*(1), 59–69.
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, *29*(1), 98–104.
- Opolko, F., & Wapnick, J. (1989). McGill University master samples user's manual. Montreal, Canada: McGill University, Faculty of Music.
- R Development Core Team. (2016). *R: A language and environment for statistical computing [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org/>
- Sjerps, M. J., Zhang, C., & Peng, G. (2018). Lexical tone is perceived relative to locally surrounding context, vowel quality to preceding context. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(6), 914–924.
- Stilp, C. E. (2019a). Auditory enhancement and spectral contrast effects in speech perception. *The Journal of the Acoustical Society of America*, *146*(2), 1503–1517.
- Stilp, C. E. (2019b). Acoustic context effects in speech perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1–18. doi: <https://doi.org/10.1002/wcs.1517>
- 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. *Attention, Perception, and Psychophysics*, *72*(2), 470–480.
- Stilp, C. E., Anderson, P. W., & Winn, M. B. (2015). Predicting contrast effects following reliable spectral properties in speech perception. *The Journal of the Acoustical Society of America*, *137*(6), 3466–3476.
- Stilp, C. E., & Assgari, A. A. (2017). Consonant categorization exhibits a graded influence of surrounding spectral context. *Journal of the Acoustical Society of America*, *141*(2), EL153–EL158.
- Stilp, C. E., & Assgari, A. A. (2019). Natural speech statistics shift phoneme categorization. *Attention, Perception, & Psychophysics*, *81*(6), 2037–2052.
- Troje, N. F., Sadr, J., Geyer, H., & Nakayama, K. (2006). Adaptation aftereffects in the perception of gender from biological motion. *Journal of Vision*, *6*(8), 850–857.
- Viemeister, N. F. (1980). Adaptation of masking. In G. V. D. Brink & F. A. Bilsen (Eds.), *Psychophysical, physiological and behavioural studies in hearing* (pp. 190–198). Delft: University Press.
- von Békésy, G. (1967). *Sensory perception*. Princeton, NJ: Princeton University Press.
- Warren, R. M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review*, *92*(4), 574–584.
- Watkins, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America*, *90*(6), 2942–2955.
- Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B*, *366*, 1702–1725.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.