Timbral Shepard-illusion reveals ambiguity and context sensitivity of brightness perception

Kai Siedenburg
Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4all, Carl von Ossietzky University, Oldenburg, Germany
kai.siedenburg@uni-oldenburg.de

Abstract: Recent research has described strong effects of prior context on the perception of ambiguous pitch shifts of Shepard tones [Chambers, Akram, Adam, Pelofi, Sahani, Shamma, and Pressnitzer (**2017**). Nat. Commun. **8**, 15027]. Here, similar effects are demonstrated for brightness shift judgments of harmonic complexes with cyclic spectral envelope components and fixed fundamental frequency. It is shown that frequency shifts of the envelopes are perceived as systematic shifts of brightness. Analogous to the work of Chambers *et al.*, the perceptual ambiguity of half-octave shifts resolves with the presentation of prior context tones. These results constitute a context effect for the perceptual processing of spectral envelope shifts and indicate so-far unknown commonalities between pitch and timbre perception.

[BLM]
Date Received: November 25, 2017 **Date Accepted:** January 17, 2018

1. Introduction

Since their introduction more than 50 years ago (Shepard, 1964), Shepard tones have assumed a central role in the study of pitch perception. With octave-space partials and a Gaussian spectral envelope, Shepard tones are emblematic of the distinction between pitch class and pitch height. Shifting partials along log-frequency scale while keeping the envelope fixed generates the well-known Shepard scale illusion—a sequence of tones with continually ascending or descending pitch. Another hallmark feature concerns the aspect of perceptual ambiguity. As for half-octave pitch shifts, listeners are on average equally likely to report upward or downward pitch shifts (Shepard, 1964). This effect appears to be subject to strong idiosyncratic biases, which itself has become the subject of a large body of work (e.g., Deutsch, 2013; Repp, 1997).

Recent studies scrutinized the role of prior auditory context in the processing of Shepard tone pairs (Repp and Thompson, 2010; Chambers and Pressnitzer, 2014). Lately, Chambers *et al.* (2017) capitalized on the ambiguity of half-octave Shepard tone shifts (replicated in their first experiment) and demonstrated (in a second, in addition to four other experiments) that the presence of a prior context robustly modulates listeners' pitch shift judgments. With sequences of tones from restricted pitch ranges preceding the target pair, participants tended to report shifts (up/down) that encompassed the frequency components of the context. This effect turned out to be robust across a variety of stimulus manipulations and was interpreted to reflect the enduring influence of auditory short-term memory on present auditory processing.

In this letter, I present evidence that the described effect extends to an aspect of timbre perception, namely, to shift judgments of timbral brightness. Technically, brightness constitutes a major perceptual attribute of the multifaceted notion of timbre (Siedenburg and McAdams, 2017) and its acoustical correlates are typically sought in the spectral envelope of sounds' steady-state portions. Intuitively speaking, a trumpet usually sounds brighter than a trombone, even if both instruments are played at the same pitch. It has been shown that the perceived brightness of sounds correlates with the spectral centroid, i.e., the centre of gravity of the spectrum (e.g., Schubert and Wolfe, 2006), with brighter sounds exhibiting higher spectral centroids. Yet, questions regarding the most perceptually relevant acoustical descriptors of the spectral envelope prevail (cf. Siedenburg *et al.*, 2016; Almeida *et al.*, 2017).

Although timbral brightness and pitch are usually studied in separation, core commonalities have been registered. For instance, McDermott *et al.* (2008) showed that listeners recognized transposed timbral brightness (and loudness) patterns in similar ways compared to pitch patterns, which indicates that relative representations, commonly thought of as a hallmark feature of pitch, could be a general feature of the

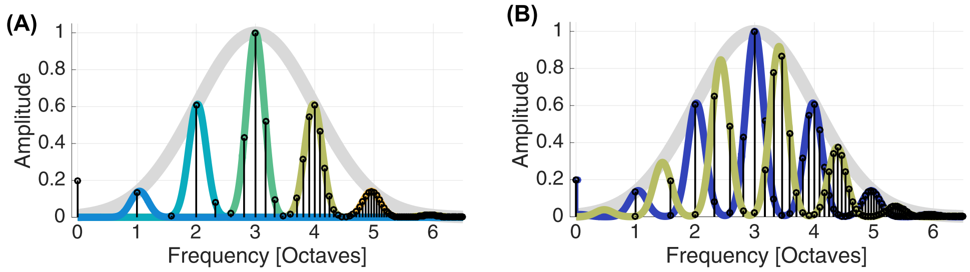

Fig. 1. (Color online) Sketches of the stimulus structure. (A) Magnitude spectra (stems) and resulting spectral envelope (colored line) and the global envelope (thick gray line). (B) Spectral envelopes separated by a half-octave shift (in dark blue and light green).

auditory system. Also testing auditory sequence discrimination, Cousineau *et al.* (2013) found strong commonalities between pitch and brightness processing for sequences of varying length, but distinct patterns of results for loudness. Allen and Oxenham (2014) studied interactions between auditory attributes. They observed symmetric mutual interference of pitch and brightness in a discrimination task, when differences in sensitivity between attributes and participants were controlled (cf. Russo and Thompson, 2005; Melara and Marks, 1990). At the same time, a recent neuroimaging study suggested no systematic anatomical distinction between the cortical regions that subserve the encoding of pitch or brightness variation (Allen *et al.*, 2016). However, to my knowledge, no reports of systematic ambiguity or context sensitivity of timbral brightness perception exist that would parallel the described effects for pitch.

To create such brightness effects, stimuli were created that conceptually resemble classic Shepard tones, but allowed to vary the spectral envelope instead of the spectral fine structure, and hence manipulate brightness instead of pitch. In contrast to the octave-spaced partials of classic Shepard tones, here sounds were characterized by harmonically spaced partials that ensured a fixed pitch percept at the fundamental frequency. The partials' amplitudes were determined by a spectral envelope that was a combination of a global bell-shaped curve and a more fine-grained cyclic envelope component that affects brightness if shifted along frequency (see Fig. 1). The resulting cyclic nature of fine-grained envelope shifts may create a timbral analogy to the classic pitch-based Shepard illusion: sound sequences with fixed pitch but continually increasing or decreasing brightness.

The experiments described in the following are modelled according to the first two experiments by Chambers *et al.* (2017). Instead of using Shepard tones or random spectra, the fundamental frequency was held constant and stimuli were varied in terms of the spectral envelopes. Experiment 1 established a relation between the magnitude of the spectral shift and listeners' perception of brightness change. Experiment 2 tested the extent to which responses for the most ambiguous shifts were modulated by prior context. Following the reported commonalities between pitch and timbral brightness perception, I hypothesized that analogous effects compared to Chambers *et al.* would unfold for brightness judgements in both experiments. Specifically, upward spectral shifts were presumed to yield brighter sounds. Given that the present stimuli shared the cyclic nature of the classic Shepard tones, the same ambiguity effects known from pitch were expected for half-octave shifts in experiment 1. In analogy with Chambers *et al.*, this ambiguity was expected to resolve with the presentation of context tones in experiment 2.

2. Methods

2.1 Participants

Participants were 18 self-reported normal hearing listeners. Three participants failed the screening test described below. The 15 participants who continued in the main experiments (11 female, 4 male, 1 other) were on average 25.8 years old ($SD = 5.2$, range: 21–43). Two participants did not have any experience in playing a musical instrument or formal music-theoretical training. The other participants reported experience in playing at least one musical instrument for 9.9 years on average ($SD = 5.4$, range: 2–18). There were two music students, and otherwise participants were amateur musicians. Participants received a compensation of 10 EUR per hour.

2.2 Stimuli

The experiment used harmonic tone complexes that were synthesized with random initial phases. On every trial, the fundamental frequency f_0 was chosen randomly between 64 and 128 Hz and partials covered the range up to 16 kHz. A global spectral envelope E was defined as range-normalized Gaussian in log-frequency with a centre of gravity three octaves above the fundamental, i.e., $\mu_E = f_0 + 3$, and a standard deviation of one octave, $\sigma_E = 1$. A more fine-grained periodic envelope component was obtained by multiplying the global envelope E with superpositions of octave-shifted, range-normalized Gaussians e with narrow shape, here chosen as $\sigma_e = 0.15$. This means, the resulting spectral envelope was the product of a wide Gaussian envelope with an octave-periodic comb, each tooth of which was a narrow Gaussian itself. Figure 1(A) shows an illustration.

Independent of the described envelope structure, the fundamental's relative amplitude was set to 0.2 to ensure a robust pitch percept at f_0 (although a clear virtual pitch at f_0 was already present due to the harmonic series). The stimulus duration was 125 ms including 5 ms raised cosine ramps at beginning and end, and the inter-stimulus interval between the reference and comparison tones was 125 ms. Sound examples are included as supplementary material.[1]

To create spectral shifts for tones with a fixed fundamental frequency f_0 and a fixed global envelope E, the fine-grained envelope components e were shifted along frequency. Figure 1(B) shows the spectral envelope of a reference tone (dark blue) together with the envelope of a comparison tone for which the components e were shifted by half an octave (light green).

Inter-trial sequences were used to minimize carry-over effects (cf. Chambers *et al.*, 2017) and consisted of sounds with a duration of 125 ms (including 5 ms raised cosine fades), separated by 125 ms inter-stimulus-intervals. All sounds had random spectra with 30 partial tones of random initial phase and random frequencies drawn uniformly (on a linear scale) between 50 and 10 000 Hz. These tone complexes were synthesized with the same spectral envelope E that was used for the target tones in the subsequent trial.

2.3 Procedure and design

Participants were instructed to listen to pairs or sequences of sounds. Their task was to judge whether the last tone was brighter (in German, "heller") than the second-to-last tone. To ensure that they understood the task, participants first completed a screening task of 40 trials comprising randomly ordered shifts with step sizes of 1/6 and 1/4 octaves. After every trial, the correct response was provided. Only participants with at least 80% correct responses were allowed to complete the main experiments (cf. Chambers *et al.*, 2017). The mean proportion of correct responses in the screening task was 0.87 (median $m = 0.93$, $SD = 0.17$, range: 0.37–1.0), and 15 out of the 18 participants scored above the criterion. The latter group had mean scores of 0.93 (median $m = 0.95$, $SD = 0.075$, range: 0.8–1.0).

In experiment 1, participants were presented with two target tones per trial, T1 and T2, and were asked to judge which of the two sounds was heard as brighter. It was randomly determined whether T1 was the reference or the shifted comparison tone. The size of the spectral shift between T1 and T2 was the independent experimental variable, and shifts by multiples of one-sixth of an octave were used (1/6, 1/3, 1/2, 2/3, and 5/6). Trials comprised an inter-trial sequence of three tones, followed by one second of silence, followed by the target stimuli T1 and T2, and a response period of variable length. No feedback about the correct response was given. Both experiments 1 and 2 comprised 40 trials per experimental condition, and the order of trials was fully randomized. The completion of experiment 1 took around 25 min.

After a break of at least five minutes, participants continued with experiment 2. Here, the target tones T1 and T2 were always separated by half-octave shifts but were preceded by a context sequence. The length of this context sequence was the independent experimental variable and varied between 0, 1, 2, and 4 tones. In half of the trials, the context tones' envelopes were shifted by intervals drawn without replacement from the set {1/12, 2/12, 3/12, 4/12, 5/12}; otherwise, shifts were drawn from the other half of the octave. Context tones otherwise had the same structure as the target stimuli. In experiment 2, the inter-trial sequence contained ten tones that had the same structure as the inter-trial tones from experiment 1. Subsequent to the inter-trial sequence that was followed by one second of silence, each trial consisted of the context tones, 250 ms of silence, the target tones T1 and T2, and a response period. The completion of experiment 2 took around 30 min.

2.4 Apparatus

Listeners were tested individually in a sound-proof lab and provided responses on a computer keyboard. Sounds were synthesized in MATLAB and were presented with an RME FIREFACE audio interface at an audio sampling frequency of 44.1 kHz and 24 bit resolution. They were presented diotically over Sennheiser HDA 200 headphones at 65 dBA sound pressure level, as calibrated by a Norsonic Nor140 sound-level meter with a G.R.A.S. IEC 60711 artificial ear to which the headphones were coupled.

3. Results

Both experiments revealed patterns of results that were very similar to the reports by Chambers *et al.* (2017). Figure 2 shows individual and mean response profiles. As expected, in experiment 1 the mean proportion of "T1 brighter" responses increased monotonically for increasing shift magnitudes. As in the case for classic Shepard tones, half-octave shifts were perceived as most ambiguous with a median of $m = 0.58$ (mean $M = 0.60$, $IQR = 0.12$). Because the data violated normality assumptions, nonparametric tests were used. A Friedman test confirmed a main effect of shift, $\chi^2(4) = 55.1$, $p < 0.001$. *Post hoc* Wilcoxon signed rank tests (two-tailed) with Bonferroni-Holm-correction indicated a monotonic increase of scores for increasing shift size (all corrected $p < 0.041$).

Following Chambers *et al.* (2017), the proportion of responses that indicated a shift in the direction of the context tones was used as the dependent variable in experiment 2. Figure 2(B) shows this variable as a function of the number of context tones. For contextless trials, responses were not biased in one particular direction (up/down), as indicated by a median of $m = 0.50$ ($M = 0.49$ and $IQR = 0.07$). Yet with four context tones, the proportion of biased responses increased up to a median of $m = 0.90$ ($M = 0.87$, $IQR = 0.15$). The overall effect of the number of context tones was confirmed by a significant Friedman test, $\chi^2(3) = 32.4$, $p < 0.001$. *Post hoc* Wilcoxon signed rank tests (two-tailed) with Bonferroni-Holm correction indicated a monotonic increase of scores for increasing number of context tones (all corrected $p < 0.02$).

Three participants were hardly biased by context with mean scores below 0.75 at four context tones. They had passed the 0.8 threshold of the screening test, with scores between 0.85 and 0.875, and were hence below the average screening score of 0.93, but did not show abnormal performance in experiment 1. This group consisted of one music student and two participants without any prior formal musical training. These outliers could be due to attention-related factors, but warrant further inquiries into inter-individual differences (cf. Pelofi *et al.*, 2017).

Whereas these experiments did not formally prove the circularity of brightness perception for tones with fixed pitch, this phenomenon is indeed inherent to the stimulus design. Sequences of stepwise shifts yield ascending or descending sequences without stark perceptual discontinuities and thus share the central characteristic of the classic Shepard illusion (Shepard, 1964). In fact, the ordinary Shepard spectrum with octave-spaced partials may even be considered as a special case of the spectra utilized here, if the width of the fine-grained (octave-periodic) envelopes σ_e is sufficiently small.

4. Conclusion

In this letter, I have described two experiments that examined timbral brightness shift judgments for tone pairs with fixed pitch, using stimuli with a harmonic spectral fine structure and cyclic spectral envelope components. The general design of the experiments followed the approach described by Chambers *et al.* (2017) and effects unfolded

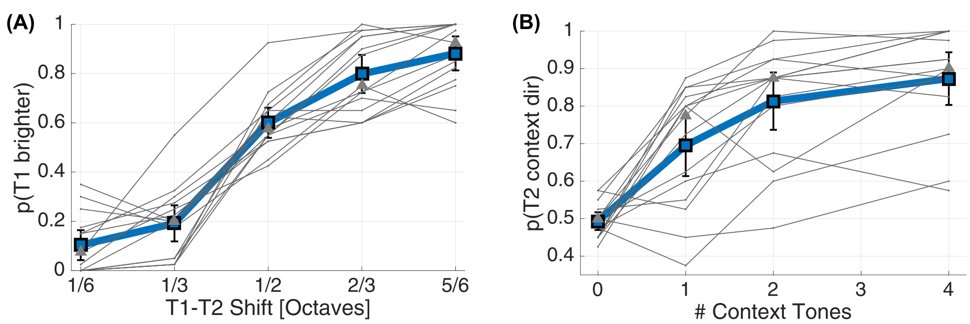

Fig. 2. (Color online) (A) Experiment 1: Proportion of "T2 lower" responses as function of shift size. (B) Experiment 2: Proportion of responses (for half-octave shifts) that indicated a shift in the direction of the context, p(T2 context dir), as a function of the number of context tones. Plots show mean scores (thick blue lines), medians (gray triangles), and individual results (thin gray lines). Error bars: 95% CI.

analogously to the original pitch experiments. Experiment 1 demonstrated that shifts of the cyclic envelope component were perceived as systematic changes in brightness. Further, half-octave shifts of the spectral envelope yielded perceptual ambiguity in the sense that participants were roughly as likely to report either one of the two presented tones as sounding brighter. Results from experiment 2 demonstrated that brightness judgments for ambiguous shifts were strongly biased towards the direction of the prior context. This finding constitutes the first systematic effect of ambiguity and context sensitivity of timbral brightness judgements. The present approach thus provides a novel window into the perceptual processing of spectral envelopes that invites further exploration.

An important open question concerns the exact nature of the cues that were most pivotal to listeners' responses in the present experiments. Although the tones' fundamental frequency did not differ within trials, it may be disputable whether the observed responses were solely based on generic timbre cues such as the energy of spectral bands of unresolved harmonics. An alternative account could state that participants used information extracted from the tracking of frequency-shifts across successive harmonics (induced by the experiments' spectral envelope shifts). In the realm of pitch, McPherson and McDermott (2018) recently provided compelling evidence in favour of this view for a broad range of classic pitch perception tasks.

From a more general perspective, the present findings provoke a comparison of pitch and timbre perception, where recent work has highlighted some critical commonalities in terms of perception and neural representation (McDermott *et al.*, 2008; Cousineau *et al.,* 2013; Allen and Oxenham, 2014; Allen *et al.*, 2016). The present findings complement this thread by indicating a shared dependency of pitch and timbral brightness on prior auditory context in situations of perceptual ambiguity. Finally, we should not forget that ideas concerning the interdependence of pitch and timbre have a long history that is not limited to psychoacoustics. As the composer Arnold Schoenberg once put it, "pitch is nothing but tone color measured in one direction" (Schoenberg, 1978, p. 421).

Acknowledgments

The author wishes to thank David Sears, Charalampos Saitis, and the reviewers for valuable comments on the manuscript. The author is also thankful to Saskia Röttges for data collection. This work was supported by a Carl von Ossietzky Young Researchers' Fellowship from the University of Oldenburg.

References and links

[1]See supplementary material at https://doi.org/10.1121/1.5022983 for sound examples.

Allen, E. J., Burton, P. C., Olman, C. A., and Oxenham, A. J. (**2016**). "Representations of pitch and timbre variation in human auditory cortex," J. Neurosci. **37**(5), 1284–1293.

Allen, E. J., and Oxenham, A. J. (**2014**). "Symmetric interactions and interference between pitch and timbre," J. Acoust. Soc. Am. **135**(3), 1371–1379.

Almeida, A., Schubert, E., Smith, J., and Wolfe, J. (**2017**). "Brightness scaling of periodic tones," Atten. Percept. Psychophys. **79**(7), 1892–1896.

Chambers, C., Akram, S., Adam, V., Pelofi, C., Sahani, M., Shamma, S., and Pressnitzer, D. (**2017**). "Prior context in audition informs binding and shapes simple features," Nat. Commun. **8**, 15027.

Chambers, C., and Pressnitzer, D. (**2014**). "Perceptual hysteresis in the judgment of auditory pitch shift," Atten. Percept. Psychophys. **76**(5), 1271–1279.

Cousineau, M., Carcagno, S., Demany, L., and Pressnitzer, D. (**2013**). "What is a melody? on the relationship between pitch and brightness of timbre," Front. Syst. Neurosci. **7**(127), 127.

Deutsch, D. (**2013**). "The processing of pitch combinations," in *The Psychology of Music*, 3rd ed., edited by D. Deutsch (Academic Press, San Diego, CA), pp. 249–325.

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2008**). "Is relative pitch specific to pitch?," Psychol. Sci. **19**(12), 1263–1271.

McPherson, M. J., and McDermott, J. H. (**2018**). "Diversity in pitch perception revealed by task dependence," Nat. Human Behav. **2**, 52–66.

Melara, R. D., and Marks, L. E. (**1990**). "Interaction among auditory dimensions: Timbre, pitch, and loudness," Percept. Psychophys. **48**(2), 169–178.

Pelofi, C., de Gardelle, V., Egré, P., and Pressnitzer, D. (**2017**). "Interindividual variability in auditory scene analysis revealed by confidence judgements," Philos. Trans. R. Soc. B **372**(1714), 20160107.

Repp, B. H. (**1997**). "Spectral envelope and context effects in the tritone paradox," Percept. **26**(5), 645–665.

Repp, B. H., and Thompson, J. M. (**2010**). "Context sensitivity and invariance in perception of octave-ambiguous tones," Psychol. Res. **74**(5), 437–456.

Russo, F., and Thompson, W. (**2005**). "An interval size illusion: The influence of timbre on the perceived size of melodic intervals," Atten. Percept. Psychophys. **67**(4), 559–568.

Schoenberg, A. (**1978**). *Theory of Harmony* (University of California Press, Berkeley, CA) (translated by R. E. Carter from original German edition, 1911, Universaledition, Vienna, Austria).

Schubert, E., and Wolfe, J. (**2006**). "Does timbral brightness scale with frequency and spectral centroid?," Acta Acust. Acust. **92**(5), 820–825.

Shepard, R. N. (**1964**). "Circularity in judgments of relative pitch," J. Acoust. Soc. Am. **36**(12), 2346–2353.

Siedenburg, K., Fujinaga, I., and McAdams, S. (**2016**). "A comparison of approaches to timbre descriptors in music information retrieval and music psychology," J. New Music Res. **45**(1), 27–41.

Siedenburg, K., and McAdams, S. (**2017**). "Four distinctions for the auditory 'wastebasket' of timbre," Front. Psychol. **8**, 1747.