Memory for Musical Key Distinguished by Timbre

Graeme Noble[1†], Joanna Spyra[1] and Matthew Woolhouse[2]
[1] Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Ontario, Canada
[2] School of the Arts, McMaster University, Hamilton, Ontario, Canada
[†] Corresponding author: graemesnoble@gmail.com

Introduction

Previous research has investigated the extent to which an initial key or pitch information is retained in short-term memory (e.g., Krumhansl & Iverson, 1992). Woolhouse, Horton & Cross (2016) empirically estimated the effect of the initial key to last between 10–12 seconds following modulation; Farbood (2016) extended this retention period to roughly 20 seconds. In a subsequent study, Spyra, Stodolak & Woolhouse (2019) evaluated whether the duration or number of chords constituting the newly modulated key was responsible for memory decays of the initial key. Their findings showed that the duration of the second, modulated key was the critical factor, rather than the number of chords. In a similar vein, recent research suggests that surface features enhance retention of nonadjacent keys in short-term memory when we listen to modulating musical sequences. For example, Spyra & Woolhouse (2018) found that the addition of figuration (e.g., passing tones and suspensions) and rhythmical activity (i.e., note density) enhanced memory for an initial key after modulation.

The studies referred to above assessed the effect of the initial key via a probe chord or probe cadence paradigm. Simply put, participants were asked to rate the probe segment's goodness-of-completion with respect to the preceding musical sequence, where the probe had either a tonic- or non-tonic-key relationship to the initial key (Farbood, 2016, used harmonic tension ratings). This paradigm has been repeatedly shown to provide a robust subjective measure of nonadjacent key effects and is assumed to reflect the retention of keys in short-term memory post-modulation. While the above research has covered several distinct aspects of music (e.g., number of chords, figuration), the contribution of timbre in this regard has yet to be studied. The research reported here addresses this lacuna.

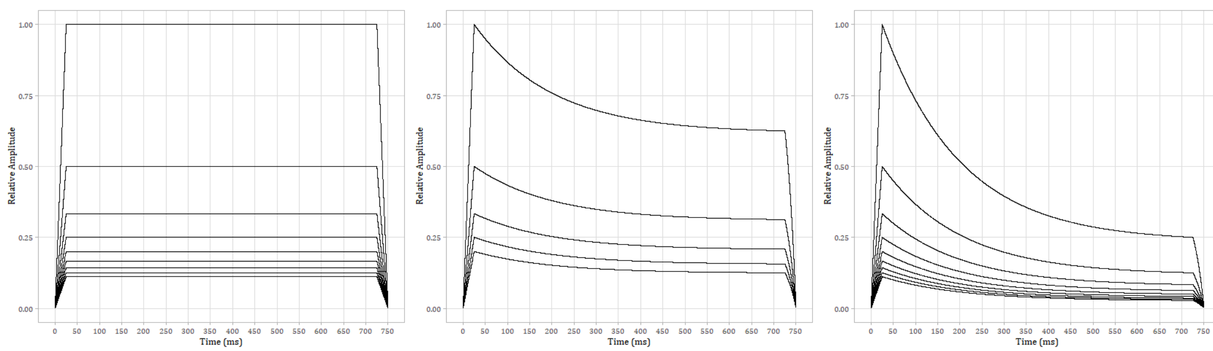
Our study operationalized timbre as consisting of pure-tone components with independent amplitude envelopes (AEs) varying in duration and intensity. The study explored these two aspects of timbre with respect to number of harmonics and AE shape. To validate an underlying assumption of the study—that artificial qualities of sounds are distinguishable—participants in Experiment 1 were asked to distinguish between three timbres constructed with varying degrees of 'naturalness'. In Experiment 2, the sounds from Experiment 1 were used to address whether timbral naturalness impacted key memorization. Using the sequences from Experiment 2, Experiment 3 explored the extent to which individual components of timbre influenced key memorization. Three hypotheses underpinned the study: (1) that timbres are distinguishable in terms of naturalness (Experiment 1); (2) that matching timbres between nonadjacent sections elicit higher goodness-of-completion ratings (Experiment 2); and (3) that distinct elements of timbre, as operationalized above, elicit different ratings (Experiment 3).

Method

Participants in Experiment 1 (n = 64; gender: 14 male, 48 female, 2 non-binary; age: range [*R*] = 17–37, mean [*M*] = 18.38, standard deviation [*SD*] = 2.43) were asked to rate the naturalness of sampled vs. subtractive vs. synthesized sounds. Typically, sounds in nature have a complex mix of harmonic and inharmonic components with independent AEs. Similarly, sampled sounds of acoustic instruments, in our case a grand piano, possess multiple harmonics with independent AEs, many of which have non-integer ratio relationships (Deutsch, 2013). The subtractive timbre was constructed as a fast Fourier transform of a sampled sound, in which inharmonic and upper harmonic elements were omitted. In our case, the subtractive timbre functioned as an intermediate between juxtaposed sampled and additively synthesized timbres. The synthesized sound was comprised of a single sine-tone played with a flat AE.

In Experiment 2 (n = 60; gender: 12 male, 46 female, 2 non-binary; age: R = 17–50, M = 18.97, SD = 4.78), three timbres—as described above—were systematically distributed across three distinct major-key stimulus segments, using the nonadjacency paradigm: (1) an initial 8-chord sequence that complied with voice-leading principles of the Common Practice period (c. 1600–1900); (2) an intervening, 8-chord modulated key sequence; and (3) a three-chord probe cadence (ii-V-I), nonadjacent to the initial key. The tempo of the stimuli was 80 BPM; each chord was 750 ms in duration. The overall duration of the nonadjacent section was therefore 6 s, followed by a 6 s intervening section, 0.75 s rest, and 2.25 s probe cadence. The timbre of the initial segment and probe cadence either matched or differed from each other, with timbres consisting of sampled audio (i.e., most natural; complex), and/or a sine tone with a flat envelope (i.e., least natural; simple). The intervening sequence consistently used the subtractive timbre.

Using the nonadjacency paradigm, in Experiment 3 (n = 33; gender: 6 male, 27 female; age: R = 17–19, M = 18.24, SD = 0.55), AEs and the number of harmonics were manipulated independently such that the initial segment and probe cadence were played with flat or dynamic AEs and either 1 or 9 harmonics. The intervening sequence had a timbre of intermediate complexity (5 harmonics with quasi-dynamic AEs; H_5–AE_{Int}; see Fig. 1b). As in previous studies that employ this paradigm, participants in Experiments 2 and 3 provided goodness-of-completion ratings for the probe cadence using a 7-point Likert-type sliding scale.



Figures 1a, 1b, and 1c: H_9–AE_{Flat}, H_5–AE_{Int}, and H_9–AE_{Dyn} timbres from Experiment 3.

Analysis

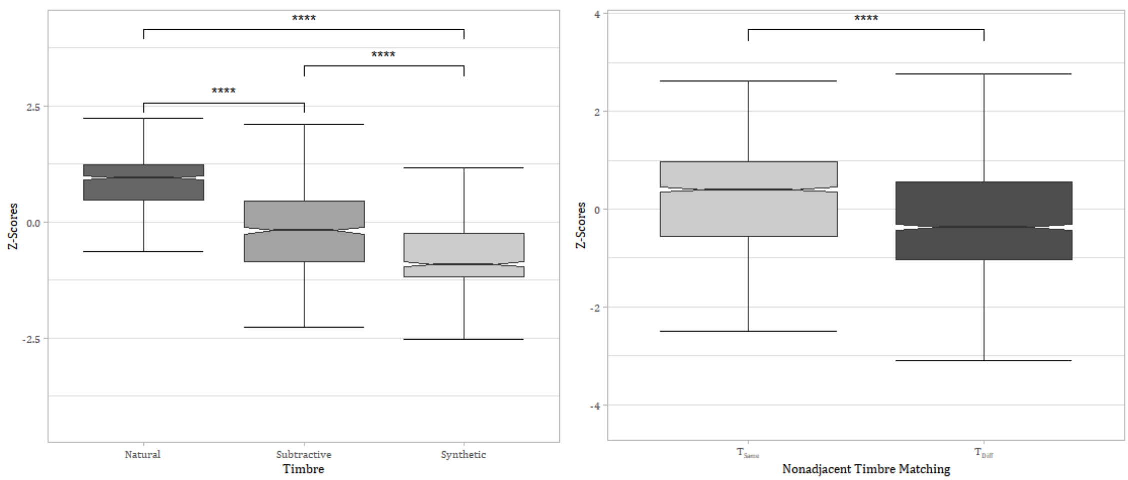
Rating data were normalized using each participant's mean and standard deviation; this process was applied to each experiment. In Experiment 1, we used a two-factor within-subjects ANOVA with *Timbre* and *Pitch Height*. *Timbre* had three levels: sampled, subtractive, and synthesized. *Pitch Height* had 12 levels, the 12-semitone transpositions of the stimuli from F3–F4 (i.e., 175 Hz to 349 Hz). Experiment 2 was assessed with a one-factor within-subjects ANOVA with *Timbre Matching* (T_{Same} vs. T_{Diff}). For Experiment 3, a three-factor within-subjects ANOVA was conducted with *Probe Timbre, Timbral-Component Matching*, and *Key Relationship*. *Probe Timbre* had four levels produced through a combination of 1 or 9 harmonics (H) and flat or dynamic AEs: H_1–AE_{Flat}; H_1–AE_{Dyn}; H_9–AE_{Flat} (see Fig. 1a); H_9–AE_{Dyn} (see Fig. 1c). Using multiple pair-wise comparisons (Tukey HSD), *Timbral-Component Matching* isolated the effects of harmonics vs. AEs. The four levels within this factor were produced through a combination of same or different number of harmonics and same or different AEs between the initial section and probe cadence: H_{Same}–AE_{Same}; H_{Same}–AE_{Diff}; H_{Diff}–AE_{Same}; H_{Diff}–AE_{Diff}. *Key Relationship* had two levels: either tonic or non-tonic between the initial section and probe cadence.

Results

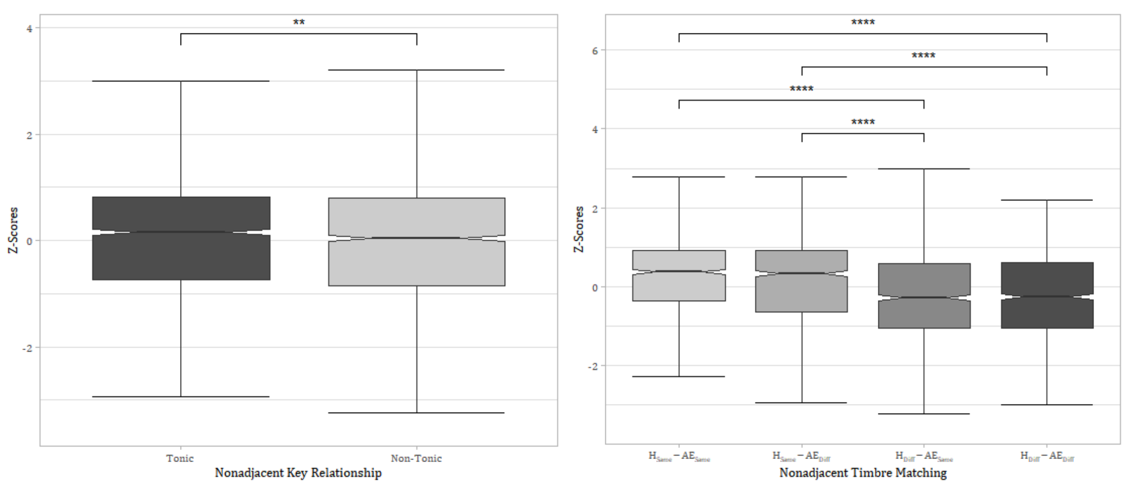
Data from Experiment 1 were consistent with the hypothesis: higher ratings of naturalness were obtained for sampled audio over subtractive over additive synthesized sounds ($F_{2, 64}$ = 776.23, p < .0001, ω^2p = .068); see Fig. 2a. These observations were consistent across pitch height transformations. In Experiment 2, participants' judgement of goodness-of-completion was greatest when the probe cadence used a timbre that matched that of the nonadjacent segment ($F_{1, 60}$ = 121.8, p < .0001, ω^2p = .039); see Fig. 2b.

Results from Experiment 3 revealed a significant effect of tonic-key relationship between the initial segment and probe cadence, independent of timbral manipulation, ($F_{1,\,33} = 7.85$, $p = .005$, $\omega^2 p = .002$); see Fig. 3a. Participants' judgement of goodness-of-completion was greatest when the probe cadence used the most natural timbre ($F_{3,\,33} = 27.00$, $p < .0001$, $\omega^2 p = .018$) or a matching timbre to the initial segment, ($F_{3,\,33} = 77.50$, $p < .0001$, $\omega^2 p = .051$); see Fig. 3b. There were no significant interactions.

Pairwise comparisons for *Timbral-Component Matching* revealed a significant contribution of number of harmonics, but not of AE. The rationale for this deduction is as follows: there was no significant difference between H_{Same} conditions ($p > .1$), whereas there was a significant difference between AE_{Same} conditions ($p < .0001$), suggesting that number of harmonics between nonadjacent sections led to consistent ratings whereas AE matching did not. We thus conclude that the harmonic component of timbre within our stimuli was critical with respect to observed nonadjacent effects.



Figures 2a and 2b: Boxplots of Experiment 1 timbre and Experiment 2 timbre matching.



Figures 3a and 3b: Boxplots of relationship and timbre matching, Experiment 3.

Discussion

The results of all three experiments demonstrate the importance of timbre for the perception of musical form involving discrete, nonadjacent sections. Once the groundwork was laid in Experiment 1, by which natural vs. synthetic sounds were shown as easily distinguishable, Experiment 2 conveyed the significance of timbral matching between nonadjacent sections within a piece of music. Irrespective of whether the nonadjacent timbres were natural or synthetic, the results showed that timbral matching across the 15 s span of each stimulus was rated higher than non-matching. Experiment 3 sought to discover whether the harmonic component or AE shape contributed to nonadjacent memory effects. Via pairwise comparisons, harmonics were found to be largely responsible for the significant variance in the data, and thus we concluded that the number of harmonics played a crucial role within our stimuli.

The results of the discrimination task in Experiment 1 somewhat echo the findings of Siedenburg and McAdams (2017), in which musicians better recognized familiar natural sounds compared to unfamiliar synthetic ones. While our task was not based on explicit recognition, it clearly showed that participants were acutely sensitive to sounds which had an acoustic origin vs. those which did not.

Data from Experiments 2 and 3 expand upon the findings of previous research using the nonadjacency paradigm (Farbood, 2016; Spyra, Stodolak, & Woolhouse, 2019; Spyra & Woolhouse, 2018; Woolhouse, Cross, & Horton, 2016). Although the 6 s intervening section used within our experiments was well below the memory limits found by the above researchers, our study is the first to investigate the influence of timbre on the perception of key nonadjacency in music. The tripartite structure of the nonadjacency paradigm is analogous to the stimuli used in Mercer and McKeown (2010), in which participants compared initial and comparison probe tones over a 10 s interval, separated by a distractor tone. The distractor varied in the number of features it shared with the initial and probe tones; performance was degraded when the distractor either consisted of novel, unshared features or contained the distinguishing feature of the probe. The critical role of the intervening distractor tone suggests that a future study should analyze the potential effects of the intervening timbre on memory for key.

The multiple comparisons conducted in Experiment 3 indicate that the harmonic content of sound is encoded in memory with greater saliency than AEs. That said, it should be noted that the comparison in Experiment 3 between number of harmonics and AEs is not a fair test. Future studies might investigate the relative influences of harmonics and AEs to the perception of timbral difference—for instance, our harmonic manipulations may have simply outweighed those chosen for AEs, resulting in imbalanced contributions to the perceived timbre. Understanding the perceptual balance between harmonic and AE manipulations is crucial for the further refinement of our experimental paradigm. One might also expand our research through alternative dependent measures (e.g., musical tension, timbre recognition or discrimination) to parse the independent effects of timbral components on music perception.

References

Deutsch, D. (ed). (2013). *Psychology of music. (pp. 26–58)*. San Diego: Elsevier Academic Press. Elsevier.

Farbood, M. (2016). Memory of a tonal center after modulation. *Music Perception*, *34*(1), 71–93. doi:10.1525/mp.2016.34.1.71

Mercer, T., & McKeown, D. (2010). Updating and feature overwriting in short-term memory for timbre. *Attention, Perception, & Psychophysics*, *72*(8), 2289–2303. doi:10.3758/BF03196702

Siedenburg, K., & McAdams, S. (2017). The role of long-term familiarity and attentional maintenance in short-term memory for timbre. *Memory*, *25*(4), 550–564. doi:10.1080/09658211.2016.1197945

Spyra, J., Stodolak, M., & Woolhouse, M. (2019). Events versus time in the perception of nonadjacent key relationships. *Musicea Scientiae*, 1–14. doi:102986491986746

Spyra, J., & Woolhouse, M.H. (2018). Effect of melody and rhythm on the perception of nonadjacent harmonic relationships. In Parncutt, R., & Sattmann, S. (eds) *Proceedings of ICMPC15/ESCOM10*, 421–425. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1131–1142. doi:10.1037/0278-7393.30.5.1131

Woolhouse, M., Cross, I., & Horton, T. (2016). Perception of nonadjacent tonic-key relationships. *Psychology of Music*, *44* (4), 802–815. doi:10.1177/0305735615593409