International Journal of Psychophysiology 129 (2018) 31–40

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Syntactic processing in music and language: Effects of interrupting auditory streams with alternating timbres

Anna Fiveash[a,b,c,e,*], William Forde Thompson[a,e], Nicholas A. Badcock[d,e], Genevieve McArthur[d,e]

[a] Department of Psychology, Macquarie University, Australia
[b] CNRS, UMR5292; INSERM, U1028; Lyon Neuroscience Research Centre, Auditory Cognition and Psychoacoustics Team, Lyon, F-69000, France
[c] University Lyon 1, Villeurbanne, F-69000, France
[d] Department of Cognitive Science, Macquarie University, Australia
[e] ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Australia

ARTICLE INFO

Keywords:
Music
Language
Syntax
Timbre
Attention
ERAN

ABSTRACT

Music and language both rely on the processing of spectral (pitch, timbre) and temporal (rhythm) information to create structure and meaning from incoming auditory streams. Behavioral results have shown that interrupting a melodic stream with unexpected changes in timbre leads to reduced syntactic processing. Such findings suggest that syntactic processing is conditional on successful streaming of incoming sequential information. The current study used event-related potentials (ERPs) to investigate whether (1) the effect of alternating timbres on syntactic processing is reflected in a reduced brain response to syntactic violations, and (2) the phenomenon is similar for music and language. Participants listened to melodies and sentences with either one timbre (piano or one voice) or three timbres (piano, guitar, and vibraphone, or three different voices). Half the stimuli contained syntactic violations: an out-of-key note in the melodies, and a phrase-structure violation in the sentences. We found smaller ERPs to syntactic violations in music in the three-timbre compared to the one-timbre condition, reflected in a reduced early right anterior negativity (ERAN). A similar but non-significant pattern was observed for language stimuli in both the early left anterior negativity (ELAN) and the left anterior negativity (LAN) ERPs. The results suggest that disruptions to auditory streaming may interfere with syntactic processing, especially for melodic sequences.

Music and language share similarities in lower-level perceptual features and higher-level structural features. Lower-level features include changes in pitch, timbre, timing, and intensity, which are fundamental characteristics of both music and language (e.g., notes, phonemes). Higher-level features emerge when smaller elements are combined through processes of auditory streaming, which form larger sequences such as musical phrases in music, or linguistic phrases in language. Musical and linguistic phrases are characterized by *syntactic structure*—a system of regularities in how elements are combined (Patel, 2008). Syntax can include hierarchical (nested) structure and strong dependencies between elements (Koelsch, 2013; Patel, 2003, 2008). Implicit knowledge of syntax results in expectations about upcoming events (Huron, 2008).

Although the elements of music and language are different (e.g., notes and chords versus words), there are parallels in how the two domains are processed in the brain. Models of music perception (Koelsch, 2011) and auditory sentence processing (Friederici, 2002) suggest similar processing stages, and it has been suggested that music and language draw upon shared resources for processing syntactic structure (Fedorenko et al., 2009; Fiveash and Pammer, 2014; Koelsch et al., 2005; Patel, 2003; Sammler et al., 2013; Steinbeis and Koelsch, 2008). The neurocognitive model of music perception (Koelsch, 2011) suggests that musical feature extraction (including pitch, timbre, and intensity information) occurs within the first 100 ms after stimulus onset. The neurocognitive model of auditory sentence processing (Friederici, 2002) also contains an early feature extraction section termed *primary acoustic analysis*—occurring within the first 100 ms after stimulus onset, before identification of word category and syntactic structure building. These early feature extraction stages feed directly into processes of auditory scene analysis.

Auditory scene analysis is the process by which incoming acoustic information is streamed into meaningful units (Bregman, 1990). The incoming information is grouped into an auditory stream based on Gestalt principles that identify the source of the sound. Sounds that are *similar* (e.g., in timbre) or *proximal* (e.g., in pitch) tend to be grouped within the same auditory stream, as they are likely to arise from the

[*] Corresponding author at: Department of Psychology, Macquarie University, Australia.
E-mail address: anna.fiveash@inserm.fr (A. Fiveash).

https://doi.org/10.1016/j.ijpsycho.2018.05.003
Received 4 May 2017; Received in revised form 5 May 2018; Accepted 7 May 2018
Available online 08 May 2018

same source (Bregman, 1990; Deutsch, 2013; Iverson, 1995; McAdams, 2013). Auditory streaming is a pre-requisite for later higher-level processes, such as developing a syntactic representation of incoming information (Koelsch, 2013). If auditory streaming were disrupted, syntactic processing should also be impaired. Given the important role of timbre in auditory scene analysis, a sequence of multiple unpredictable timbres are unlikely to be grouped as part of the same auditory stream (Bregman, 1990). This disruption to auditory streaming would in turn have an impact on syntactic processing. Thus, timbre can be used as a tool to investigate processes of auditory streaming and syntactic processing in music and language.

Links between timbre and syntax have been observed in previous research. For example, McAdams (1999) presented participants the same piece of music with either a sampled orchestra (multiple timbres) or a sampled piano (one timbre). McAdams (1999) stopped the music at 23 distinct points, and asked participants at each point to rate how "complete" the music sounded. Lower ratings of completion imply higher tonal tension, whereas higher ratings of completion suggest lower tonal tension (more relaxation). Ratings of completion were significantly higher for the orchestral version of the piece than for the piano version. As musical tension and relaxation patterns are integral to syntactic structure (Huron, 2008), it appears that participants were less sensitive to tension and resolution patterns in the syntax when they were presented with multiple instruments, suggesting that timbre influenced sensitivity to syntactic structure. Cusack and Roberts (2000) further showed that changing timbres in a rhythm discrimination task resulted in poorer performance on a task requiring stream integration, as changing timbres disrupted this process.

Recent evidence also suggests that syntactic processing is less engaged by melodic sequences that contain alternating timbres (Fiveash et al., under review). In Experiment 1, participants listened to melodies with one- and three-timbres (among other conditions) while recalling complex sentences or word-lists. Participants were better able to recall complex sentences when they were accompanied by melodies with three timbres compared to melodies with one timbre. At first glance, this finding may seem surprising since changing-state stimuli tend to be more distracting and hence would be expected to result in increased interference with sentence processing (Jones et al., 2010). However, this finding may suggest that the frequent changes in timbre interrupted the auditory streaming of melodic information, resulting in less coherent musical sequences and weaker music syntax processing, and hence less interference on linguistic syntax processing.

To examine this hypothesis, Fiveash et al. (under review) conducted a second experiment in which they asked participants to compare two sequential melodies. Participants were worse at discriminating between melodies that contained timbre changes than melodies comprised of a single timbre, indicating that changes in timbre made it difficult for listeners to form a stable and coherent mental representation of the melodies. Although previous research has revealed links between timbre and auditory streaming (Bregman, 1990), and between timbre and syntax (McAdams, 1999), this is the first study to show that interrupting an auditory stream with changes in timbre reduces syntactic processing. These findings are consistent with the possibility that timbre affects syntactic processing because of its powerful role in auditory streaming (Bregman, 1990).

Based on auditory streaming research (Bregman, 1990; Deutsch, 2013; Iverson, 1995), links between timbre and syntax (Cusack and Roberts, 2000; Fiveash et al., under review; Koelsch, 2013; McAdams, 1999), and parallels between music and language (Fiveash and Pammer, 2014; Jentschke et al., 2005; Koelsch et al., 2002; Kunert et al., 2015; Levitin and Menon, 2003; Maess et al., 2001; Masataka, 2009; Patel, 2008), we predicted that participants would be less sensitive to violations of syntactic structure in both music and language when the auditory streams were interrupted with alternating timbres. To evaluate this prediction, we used event-related potentials (ERPs) to measure brain responses to syntactic violations in normal and interrupted melody and sentence streams.

ERPs are used to measure the timing of brain responses to various stimuli (Luck, 2014). ERP studies have established that when participants hear an out-of-key note or musical chord (a violation of syntactic structure), an early right anterior negativity (ERAN) ERP component is elicited approximately 170-220 ms after stimulus onset (Koelsch, 2013). This component is measured by calculating a *difference ERP waveform* that represents the difference between the ERP to a syntactic violation in a melody and the ERP to the same point in the same melody with no such violation present. It has been suggested that the ERAN reflects an interruption to initial structure-building processes in the brain (Koelsch, 2013). The ERAN is reliably elicited to out-of-key chords in a sequence, and out-of-key notes within a melody (e.g., Koelsch et al., 2000; Koelsch et al., 2005; Koelsch and Jentschke, 2008; Miranda and Ullman, 2007).

It is important to ensure that the ERAN reflects syntactic processing and not merely the processing of deviant elements. Koelsch and colleagues have argued that the ERAN is distinct from the mismatch negativity (MMN)—an early component elicited in oddball paradigms to a physical or abstract feature deviant (Koelsch et al., 2001). One reason for this distinction is that the ERAN is affected by tonal context; that is, the amplitude is directly related to how unexpected a tone or chord is within a current key. In contrast, the MMN is not affected by tonal context. Thus, Koelsch et al. (2001) concluded that the MMN is a response to physical features of a stimulus and does not reflect sensitivity to the tonal relationships established by a musical key. The MMN is elicited even under heavy sedation whereas the ERAN is not, suggesting two distinct neural indices (Koelsch et al., 2006). Moreover, when the ERAN is evoked by violations to music syntax, concurrent violations to language syntax interact with this brain response. Such an interaction is not observed for the MMN response (Koelsch et al., 2005). The combination of these findings suggests that the ERAN component is related to music syntax processing in the brain, and distinct from the MMN.

ERPs can also be used to examine syntactic processing in language. To date, two early ERP components to syntactic violations in language have been identified. The early left anterior negativity (ELAN) has been found in response to word-category violations and early phrase-structure violations, and occurs at around the same time as the ERAN (i.e., 100–300 ms post stimulus onset; Friederici, 2002). A left anterior negativity (LAN) is evident at approximately 300–500 ms post stimulus onset, and is found with morpho-syntactic violations, number disagreements, and gender disagreements (Coulson et al., 1998; Friederici, 2002; Gunter et al., 2000). Previous research has also found the LAN in response to word-category violations (e.g., Hagoort et al., 2003). The LAN is in the same time window as the N400—a component elicited with semantic errors in language. However, research suggests that these reflect two separate processes due to a lack of interaction between the two components (Friederici, 2002; Gunter et al., 2000). Thus, it appears that the ELAN and LAN are early indicators of a syntactic violation in language, and not just a generic response to a violation of expectations.

The current study investigated syntactic processing in music and language using the ERAN, ELAN, and LAN ERP components, as well as behavioral measures. Specifically, we were interested in whether disrupting auditory streams with alternating timbres has analogous effects on neurophysiological and behavioral indices of syntax processing in the two domains. We tested this by determining, firstly, whether our stimuli elicited expected brain responses to syntactic violations (ERAN in music, and the ELAN or LAN in language). Once we identified these components, we determined whether the response to a syntactic violation was significantly reduced in the three-timbre conditions (disrupted auditory streaming) compared to the one-timbre conditions (intact auditory streaming). A reduced response in the three-timbre conditions would indicate that alternating timbres led to a reduction in syntactic processing. However, no difference would suggest that alternating timbres did not have an impact on syntactic processing at the neurophysiological level. A similar pattern in both music and language would indicate that a disruption to auditory streaming leads to a reduction in syntactic processing that operates in a similar way across both domains. The current investigation is the first to examine the electrophysiological consequences of disrupting auditory streaming with changes in timbre, and how this disruption impacts syntactic processing in music and language.

1. Method

1.1. Ethics

This study was approved by the Macquarie University Human Research Ethics Committee (ref: 5201500300).

1.2. Participants

Twenty-three students from Macquarie University participated for course credit. One participant was excluded due to a recording error, leaving 22 participants ($M_{\text{age}} = 20$ years, range: 18–24; 17 females). All were native English speakers, and 21 reported being right handed. Participants had an average of 3.89 years of private music lessons (range: 0–14), and 6.5 years of private, classroom, and self-taught musical experience (range: 0–14). Eight participants had five or more years of private music training. Three participants indicated that they were musicians, 14 indicated they were non-musicians, and four considered themselves as *somewhat* a musician (one participant did not respond to this question). Eight indicated that they were currently musically active. All reported listening to music daily, with an average of 124 min per day ($SD = 90$ mins, range: 10–360 min).

1.3. Design

The experiment consisted of separate melody and sentence blocks, both with a 2 (timbre: one, three) × 2 (syntax: violation, no-violation) within-subjects design. There were four conditions within each block, with 50 trials in each condition (i.e., a total of 200 melody trials and 200 sentence trials per participant). The melodies were played with one timbre or three timbres, with a violation (out-of-key note) or no violation. The sentences were spoken by one speaker or three speakers (voices), with a violation (phrase-structure violation) or no violation. Presentation of melody and sentence blocks was counterbalanced across participants. Within blocks, stimulus presentation was randomised to ensure different presentation for each participant. The same melody or sentence was never presented consecutively. Both behavioral and ERP data were recorded simultaneously, and participants had a break every 50 trials.

1.4. Stimuli

Stimuli were programmed and presented using Matlab (version R2016b) and Psychtoolbox (version 3.0.13, Brainard, 1997; Kleiner et al., 2007).

1.4.1. Melodies

Fifty musical instrument digital interface (MIDI) melodies were created in MuseScore in the majorkeys of C, G, D, and A. The melodies were composed by a professional composer (the second author), and simplified for ERP research by the first author. All melodies started and ended on the tonic note of the key to enhance key strength, were 100 bpm, four bars long, and in a 4/4 time signature. Melodies contained 21 notes on average (range: 18–24 notes) (see Fig. 1 for an example melody). MIDI melodies were then imported into GarageBand. One-timbre conditions were played on the Steinway grand piano MIDI instrument, and three-timbre conditions were played with Steinway grand piano, acoustic guitar, and vibraphone MIDI instruments. An external random number generator determined which instrument played each note, and it was ensured that no instrument played more than two notes in a row. Acoustic guitar, grand piano, and vibraphone instruments were chosen for three main reasons. First, they are relatively familiar to participants; second, they can all be played in the same (equal temperament) tuning system; and third, they are all characterized by a rapid attack time (e.g., the energy in the note reached its peak quickly), which minimized perceived differences in note onset times (McAdams, 2013).

The one- and three-timbre violation conditions contained an out-of-key note. Stimuli were designed so that the critical note (out-of-key note) was always in the final two bars, always fell on a strong (one or three) beat, on a full quarter note, and was always preceded by a full quarter note. This ensured that there was always 600 ms in note length to measure the violation response (i.e., the baseline was not corrupted by the onset of a previous note). Out-of-key notes were always within three semitones of the original note ($M = 1.16$ semitones, $SD = 0.51$ semitones). Where possible, notes were only changed by one semitone. This manipulation maintained the melodic contour of the melody but introduced a note that did not occur within the key. For example, a C note in C major could be altered to a C sharp ($C^{\#}$) note. There was a range of different out-of-key notes depending on the key of the melody. Out-of-key notes consisted of B flat (B^{b}; in the keys of A, C, D, and G major), C (in the key of A), $C^{\#}$ (in the keys of C and G major), $D^{\#}$ (in the keys of A, C, and D major), F (in the key of D major), $F^{\#}$ (in the key of C major), G (in the key of A), and $G^{\#}$ (in the keys of C, D, and G major).

1.4.2. Sentences

Sentences were designed for the same four conditions: One-timbre (violation, no-violation), and three-timbres (violation, no-violation). Thirty sentences from Neville et al. (1991) were used, and 20 more with a similar structure were created so there were 50 sentences in total. These sentences, each comprising seven or eight words, were all declarative sentences consisting of noun phrases and a possessor (e.g., Fred's). The sentences all had a similar structure, such as: *The widow asked for Fred's advice* ***about*** *taxes*. Phrase-structure violations were used, as these have been shown to disrupt early syntactic processing, akin to music syntax violations (Koelsch, 2013). To create the phrase structure violation, the critical word (always *about* or *of*), was moved to the position after the possessor, such as: *The widow asked for Fred's* ***about*** *advice taxes*. For more information about the sentence constructions, please see Neville et al. (1991).

To create the three distinct voice timbres for the three-timbre condition, three Australian, female, native-English speakers with clearly distinctive vocal timbres were chosen. Female voices were used for two main reasons. The first reason was to introduce clear changes in timbre without also introducing sudden changes in pitch. Dramatic changes in the pitch of speakers would have added a second source of distraction to auditory sentence processing beyond changes in timbre, and would have introduced a serious experimental confound. Thus, both music and language conditions were constructed to introduce changes in timbre but not in pitch. This decision ruled out alternations between male and female voices, for example. The second reason was based on the consideration that *expectations* for timbre shifts in speech and music may differ. Whereas vocal timbre while speaking a sentence is typically

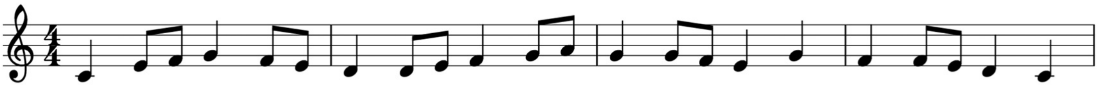

Fig. 1. Example melody in C major.

constrained by the qualities of a single speaker, it is not uncommon for novel timbres to be introduced during music listening. As such, any obvious changes in the speaker during a spoken sentence should be highly salient.

Speakers were recorded in a sound proof room. The speakers practiced before recording. They were instructed to read each sentence with normal prosody, but with gaps after each word that were long enough that the words did not run together. This manipulation allowed for word splicing (Praat, version 5.4.22; Boersma, 2001), and minimised overlap between ERPs to successive words. Praat was then used to ensure there was always at least 600 ms from the onset of one word to the onset of the next word, and at least 100 ms of silence before the onset of each word to maximize a stable pre-stimulus baseline (see below). The same speaker spoke all sentences in the one-timbre condition. To create the three-timbre condition, an external random number generator determined which speaker would speak each word (with the caveat that the same speaker never spoke two words in a row). Different speakers' voices were then spliced together using Praat to create sentences. The dynamic range of the three-timbre sentences was compressed using Audacity to ensure no large fluctuations in voice loudness.

It should be noted that the music stimuli allowed two instances of the same instrument in a row, whereas the speaker changed for each word in the sentences. This difference occurred because the melodies contained 18–24 notes, whereas the sentences only contained seven or eight words. By alternating voice timbres on each word, it was ensured that each speaker spoke at least two words in each sentence, maximising timbral variation.

1.4.3. *Critical points*

Critical points in the stimuli were marked using Praat. For the music stimuli, the critical time points were at the onset of the out-of-key note, and the onset of the same note in the matching no-violation melody stimulus. In the language stimuli, the critical time points were the onset of the violation word, and the onset of the same word in the matching no-violation sentence stimulus. Event markers were sent to the continuous EEG recording at the onset of each trial using a parallel port, and the critical time point was updated offline.

1.5. *Procedure*

Participants were tested in an electrically and acoustically shielded room. Participants signed the information and consent form, filled out a music education and preference questionnaire, and were instructed about the task. To reduce set up time by reducing electrode impedance, the participant's scalp was combed (Mahajan and McArthur, 2010), face and mastoid areas were cleaned, and electrodes were placed on the face and Mastoid bones and filled with a conductive gel. The EasyCap with electrodes was then secured on the participant's head, and scalp electrodes were filled with conductive gel. Electrode impedances (measured using the Neuroscan Synamps acquisition system and Scan software; Scan 4.3) were adjusted to be below 5 kΩ. This set-up process took approximately 30 min.

Participants were instructed that on each trial of the experiment, their task was to decide whether or not there was (1) an out-of-key note in a melody played by a piano or by three different instruments; or (2) a grammatical error in a sentence that was spoken by either one speaker or three different speakers. Participants heard examples of the stimuli. When participants indicated they understood the task and were comfortable, the experiment began. After each trial, participants indicated whether or not they detected a violation by pressing the *z* or *m* keys on the keyboard, respectively. The experiment took approximately 1 h and 30 min, including set-up time.

1.6. *Behavioral measures*

Behavioral data consisted of participant responses to the question: *was there a violation?* (yes or no) for each trial. To analyze these responses, d prime (d′) sensitivity scores and reaction times (RTs) were calculated to measure how sensitive participants were to detecting out-of-key notes in music, and grammatical errors in language. Note that participants responded to the question after the stimulus had finished, hence reaction times are more likely to reflect decision times rather than detection times. D primes were calculated and used in the analysis as a well-known and widely used measure of signal detection that allows for correction of extreme values. D prime values were calculated by subtracting the z scores for each participant's *false alarm* rate (when there was no error and the participant said there was an error) from the *hit* rate z score (when there was an error and the participant detected an error). Extreme values for hit or false alarm rates (e.g., 1 or 0) were corrected for by replacing scores of 1 with (n – 0.5), and scores of 0 with 0.5/n, as suggested in Stanislaw and Todorov (1999), where n is equal to the number of trials. A measure of response bias c was also calculated (see Stanislaw and Todorov, 1999) which revealed whether participants were more biased towards responding *yes* or *no* overall. Positive c scores reflect a bias towards responding no, and negative scores reflect a bias towards responding yes. A score of zero indicates no bias. Mean RTs for each participant were calculated across all conditions, and any RTs above three standard deviations from the mean were not included in the analysis.

1.7. *EEG recording*

Electroencephalography (EEG) was recorded using the Neuroscan system (version 4.3) and a Synamps2 amplifier with a sampling rate of 1000 Hz, and an online bandpass filter (1–100 Hz). Brain activity was measured through 30 electrodes positioned according to the 10–20 system (EasyCap; Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP8, P7, P3, Pz, P4, P8, O1, Oz, O2). The ground electrode was located at AFz, and reference electrodes were placed on the left and right mastoid bones. Horizontal electro-oculographic (HEOG) activity was recorded using electrodes placed on the left and right outer canthi of the eyes. Vertical electro-oculographic (VEOG) activity was recorded using electrodes placed above and below the left eye.

1.8. *ERP processing*

Data was processed using EEGLAB (version 13, Delorme and Makeig, 2004) and Matlab. EEG recorded from each electrode was filtered with a high-pass filter of 0.1 Hz and a low-pass filter of 30 Hz. Since the online reference was the left-mastoid (M1), the EEG data was re-referenced offline to the right-mastoid (M2). An independent components analysis (ICA) was run on all the EEG data from all electrodes in EEGLAB. Eye blink components were removed based on visual inspection of ICA components. EEG data were then epoched to 700 ms after the onset of the critical note or word, with a baseline correction of 100 ms. Epochs with extreme values at the sites of interest (frontal left and right electrodes—F7, FT7, F3, FC3, F8, FT8, F4, FC4) greater than −150, ± 150 microvolts were removed. This resulted in a 0.8% loss of epochs across the different conditions (music: one-timbre no-violation (8), one-timbre violation (12), three-timbres no-violation (7), three-timbres violation (12); language: 9, 7, 10, 7, respectively). Individual participants had between zero and 18 epochs (M = 3.2, SD = 4.6) removed (out of a possible 400). The remaining epochs in each of the eight conditions were then averaged to create ERPs of each participant's response for each condition (for both melodies and sentences: one-timbre no-violation, one-timbre violation, three-timbres no-violation, three-timbres violation).

1.9. ERP components

The ERAN component in music, and the ELAN and LAN components in language, have reliably been found at anterior sites, reflected primarily in the frontal left and right electrodes (Friederici, 2002; Koelsch, 2013; Koelsch et al., 2001; Maidhof and Koelsch, 2011). Therefore, we focused our analyses on the average of the frontal left (F7, FT7, F3, FC3) and frontal right (F8, FT8, F4, FC4) electrodes. We chose to focus our analysis on anterior sites because of our strong theoretical expectations of where the expected components would be observed. By focusing our analysis on electrodes that were determined a priori, we decreased our potential for making a Type 1 error. Further, because our stimuli contained a new event approximately every 600 ms, potential P600 effects that may have been localized in posterior regions (Friederici, 2002; Patel et al., 1998) would be influenced by the onset of the new note or word. Therefore, an anterior site analysis was most appropriate in the current case.

Considering the distinct spectral differences between the one- and three-timbre conditions, we calculated the difference waves of the violation condition minus the no-violation condition for each individual. The difference waves isolate the response to out-of-key notes in music and phrase-structure violations in language, irrespective of the sensory differences in the stimuli. Based on a visual analysis of the ERP components at both the individual and group level, and previous research (e.g., Koelsch, 2013), we defined the music ERAN time period of interest as 150–250 ms. For the language stimuli, previous research suggests that a phrase-structure violation in language results in an ELAN, reported to be around 100–300 ms (Friederici, 2002; Koelsch, 2013). However, research has also shown an anterior negativity (non-lateralized, between 300 and 500 ms) to word-category violations, suggesting that our stimuli may also elicit a later negativity (Hagoort et al., 2003). A visual analysis of our data revealed two negative going peaks in the language difference waves, which appeared to reflect both the ELAN at 100–150 ms and the LAN (Friederici, 2002; Hagoort et al., 2003; Koelsch, 2013) at 270–360 ms. We therefore analyzed both time frames in the language stimuli.

Within the time periods of interest for both music and language, we extracted the peak negativity, and calculated the 50 ms average around this peak (25 m either side) for each individual in each condition, for both hemispheres. We did this to get a sensitive measure of the brain response within the time periods of interest, tailored to each individual. For music, we extracted one peak in the ERAN time window (150–250 ms). For language, we extracted two peaks—one in the ELAN time window (100–150 ms), and one in the LAN time window (270–360 ms).

1.10. Analysis

1.10.1. Behavioral analysis

D prime values and the response bias measure c were compared in the one-timbre and three-timbre conditions separately for melodies and sentences using paired-samples *t*-tests. For the RT analysis, separate 2 (timbre: one, three) × 2 (violation: no, yes) repeated measures analyses of variance (ANOVAs) were conducted on the melody and sentence responses. To investigate significant effects, paired-samples *t*-tests were conducted with adjusted *p* values (*p*') reported (Holm-Bonferroni corrected for multiple comparisons). Years of private musical training were not correlated with d′ scores for the one-timbre ($r = 0.12$, $p = .60$) or three-timbre conditions ($r = 0.13$, $p = .57$) for melodies, or the one-timbre ($r = 0.06$, $p = .79$) or three-timbre conditions ($r = 0.12$, $p = .61$) for sentences. Therefore, participants were analyzed as one group.

1.10.2. ERP analysis

Our primary ERP analysis consisted of analyzing the hemisphere with the strongest response in each condition, as our primary goal was to investigate the difference in activation between the one-timbre and three-timbre conditions. As participants in our sample differed in their lateralization of syntactic violations for both music and language, we chose a more "individual-based" approach to answer our research question. To complement this analysis approach, and investigate the response at the group level, we also ran a secondary analysis investigating lateralization of the components. Although the ERAN is generally right lateralized (Koelsch, 2013), and the ELAN and LAN are generally left lateralized (though prosody appears to be processed in the right hemisphere; Friederici, 2002), there have also been a number of studies which have shown a bilateral distribution of both the LAN (Hagoort et al., 2003), and the ERAN (Garza Villarreal et al., 2011; Loui et al., 2005). In addition, (1) the processing of timbre in the brain is not well understood (Reiterer et al., 2008), (2) it is possible that the unusual nature of our three-timbre stimuli may have led to differences in lateralization between participants (Boucher and Bryden, 1997), and (3) differences in lateralization have also been found for musicians, who tend to show a greater bilateral distribution of the ERAN (Ono et al., 2011). These findings, combined together, suggest that the lateralization of the ERAN, ELAN, and LAN cannot be presumed in all subjects, and hence we included both an individual-level and a group-level analysis.

Our first goal was to ensure that our stimuli elicited a reliable ERAN in the melodies, and ELAN and LAN in the sentences in each participant's dominant hemisphere. One-sample *t*-tests were conducted in the time windows of interest for the one-timbre and three-timbre responses. Holm-Bonferroni adjusted *p* values (*p*') are reported for two comparisons in each time window for the sentence data.

Our second goal was to investigate the effects of alternating timbres on the ERP components related to violations of syntax in both melodies and sentences. Planned paired-samples *t*-tests were conducted on the brain responses to the one- and three-timbre conditions. These tests were based on the hemisphere with the largest response.

To complement these main analyses, and to investigate whether the components were lateralized at the group level, repeated measures ANOVAs were conducted in each time window with the factors timbre (one, three) and hemisphere (right, left). Significant effects were investigated with paired samples *t*-tests with Holm-Bonferroni adjusted *p* values reported for multiple comparisons.

Because the timbre manipulation was not directly comparable across the melody and sentence stimuli, it was not appropriate to compare the melody- and sentence-induced ERPs directly in an analysis. However, using a within-subject design allowed us to compare the effects of manipulating timbre on syntactic processing in both melodies and sentences.

2. Results

2.1. Behavioral results

2.1.1. Melodies

D prime sensitivity measures showed that participants were significantly better at detecting out-of-key notes in the one-timbre condition ($M = 2.79$, $SD = 0.45$) than the three-timbre condition ($M = 2.26$, $SD = 0.77$), $t(21) = 5.22$, $p < .001$, $d = 0.84$, see Fig. 2. Note that no participants scored 100% for out-of-key note detection in either the one-timbre (range: 48–96%) or the three-timbre (range: 54–94%) conditions. Both the one-timbre ($M = 0.62$, $SD = 0.37$) and three-timbre ($M = 0.22$, $SD = 0.38$) conditions showed a bias towards responding *no*. However, the one-timbre condition led to a significantly stronger bias towards saying no than the three-timbre condition, $t(21) = 5.29$, $p < .001$, $d = 0.62$. Therefore, it appears that participants were more biased towards reporting no error in the one-timbre condition compared to the three-timbre condition.

The RT analysis showed no main effect of timbre, $F(1, 21) = 1.44$, $p = .24$, a main effect of violation, $F(1, 21) = 4.88$, $p = .04$, $\eta^2 = 0.19$,

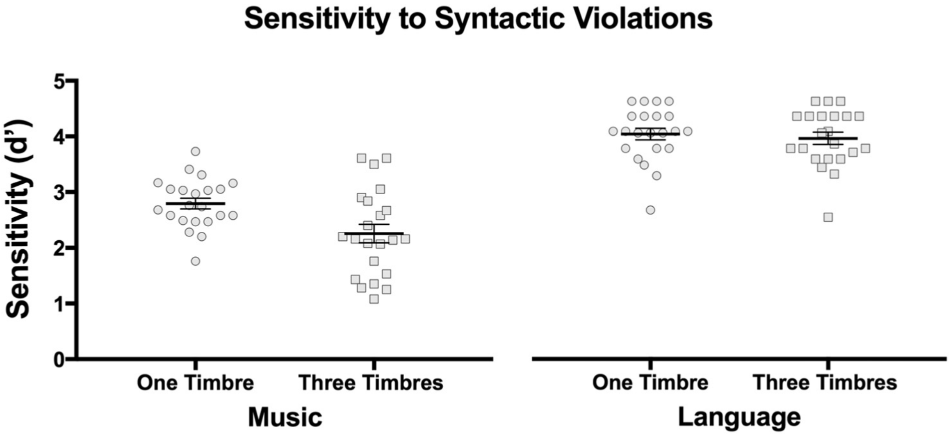

Fig. 2. D prime values reflecting sensitivity to out-of-key notes in music, and grammatical errors in language. Individual data points reflect individual participant scores, and the mean is represented by the black line. Error bars indicate one standard error either side of the mean.

and a significant interaction, $F(1, 21) = 6.11$, $p = .02$, $\eta^2 = 0.23$. Paired-samples *t*-tests revealed that participants were significantly faster at deciding whether there was a violation in the three-timbre violation condition ($M = 0.53$ s, $SD = 0.24$ s) compared to the three-timbre no-violation condition ($M = 0.63$ s, $SD = 0.27$ s), $t(21) = 3.31$, $p' = .01$, $d = 0.66$. However, there was no significant difference between the one-timbre violation ($M = 0.59$ s, $SD = 0.28$ s) and the one-timbre no-violation ($M = 0.62$ s, $SD = 0.31$ s) conditions, $t(21) = 0.99$, $p' = .66$. There were also no differences between the one-timbre and three-timbre no-violation conditions, $t(21) = 0.35$, $p' = .73$, or the one-timbre and three-timbre violation conditions, $t(21) = 2.45$, $p' = .07$. These results suggest that the main effect of violation and the interaction between timbre and violation were primarily driven by the faster RT to decide there was an error in the three-timbre violation condition.

2.1.2. Sentences

Sensitivity measures (d′) showed no difference between the one-timbre sentence condition ($M = 4.06$, $SD = 0.49$) and the three-timbre sentence condition ($M = 3.98$, $SD = 0.52$), $t(21) = 0.80$, $p = .43$ (see Fig. 2). Note that seven participants were 100% accurate in detecting violations in the one-timbre condition (range: 88–100%), and eight participants were 100% accurate in the three-timbre condition (range: 90–100%). There was also no difference between the one-timbre ($M = 0.05$, $SD = 0.19$) and three-timbre ($M = 0.05$, $SD = 0.24$) conditions in the measure of response bias c, $t(21) = 0.07$, $p = .95$. These findings may be due to ceiling effects, as the grammatical errors were very obvious, and participants detected them with high accuracy.

The RT analysis revealed a main effect of timbre, $F(1, 21) = 4.97$, $p = .04$, $\eta^2 = 0.19$, a main effect of violation, $F(1, 21) = 8.49$, $p = .01$, $\eta^2 = 0.29$, and no interaction between timbre and violation, $F(1, 21) = 0.44$, $p = .52$. When multiple comparisons were controlled for, there were no significant differences between conditions. However, there was a trend for participants to respond more quickly when there was a violation for both the one-timbre (no-violation: $M = 0.77$ s, $SD = 0.37$ s; violation: $M = 0.67$ s, $SD = 0.33$ s), $t(21) = 2.45$, $p' = .07$, and the three-timbre conditions (no-violation: $M = 0.74$ s, $SD = 0.38$ s; violation: $M = 0.62$ s, $SD = 0.30$s), $t(21) = 2.67$, $p' = .06$. The main effect of timbre occurred because there was a trend for participants to respond more quickly in the three-timbre condition ($M = 0.68$ s, $SD = 0.32$ s) compared to the one-timbre condition ($M = 0.72$ s, $SD = 0.34$ s). However, there were no significant differences when comparing the one-timbre no-violation condition with the three-timbre no-violation condition, $t(21) = 0.99$, $p' = .34$, or the one-timbre violation condition with the three-timbre violation condition, $t(21) = 2.01$, $p' = .11$.

2.2. Reliability of ERP components

2.2.1. Melodies

For the brain response to out-of-key notes in melodies, the ERAN difference wave component was statistically significantly different to zero in the ERAN time window (150-250 ms) for both the one-timbre, $t(21) = 6.66$, $p < .001$, $d = 1.42$, and three-timbre, $t(21) = 6.04$, $p < .001$, $d = 1.29$ conditions. This test confirms the existence of the ERAN to out-of-key notes in our data (see Fig. 3a and Table 1 for grand averages).

2.2.2. Sentences

In the ELAN time window (100-150 ms), the brain response to the one-timbre condition was significantly different to zero, $t(21) = 2.81$, $p' = .02$, $d = 0.60$; however, the brain response to the three-timbre condition was not, $t(21) = 1.84$, $p' = .08$. In the LAN time window (270-360 ms), the difference waves were significantly different to zero for both the one-timbre, $t(21) = 4.65$, $p' < .001$, $d = 0.99$ and three-timbre, $t(21) = 3.54$, $p' = .002$, $d = 0.75$, conditions (*p* values adjusted for two comparisons). These findings suggest that (1) the ELAN was not evident in the three-timbre condition, and (2) the LAN was evident in both the one- and three-timbre conditions.

The difference ERP wave in the LAN time window appeared more reliable than the ELAN response, due to its larger amplitude and existence in both the one- and three-timbre conditions. Thus, the following analysis focused on the LAN rather than the ELAN as a neural index of a syntactic violation in language. However, it is interesting to note that the ELAN was evident (though quite weak) for the one-timbre condition, but was not evident for the three-timbre condition. The difference between the one-timbre and three-timbre conditions in the ELAN time window was not significant, $t(21) = 1.24$, $p = .23$. See Fig. 3b for a visual representation of the LAN across all participants, and see Table 1 for grand average means and standard deviations for the ELAN and LAN.

2.3. Effect of disrupting auditory streaming

2.3.1. Melodies

Supporting our hypothesis, the ERAN to violations of music syntax was significantly more negative in the one-timbre condition compared to the three-timbre condition, $t(21) = 2.74$, $p = .01$, $d = 0.71$. Thus, the response to a music syntax violation in the three-timbre condition was reduced compared to the one-timbre condition, as predicted.

2.3.2. Sentences

The brain response to syntactic violations in sentences in the LAN

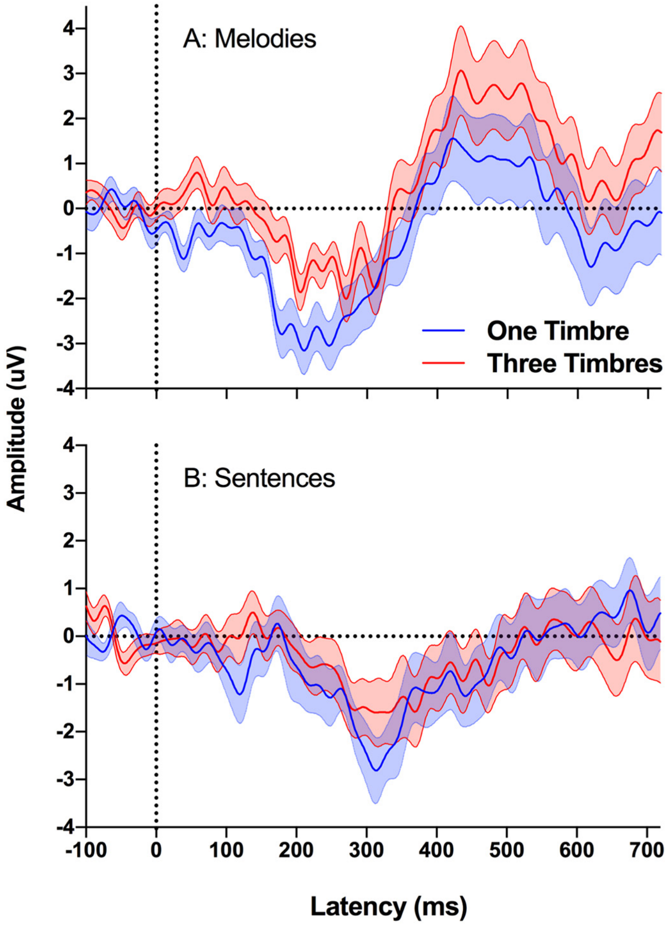

Fig. 3. Difference waves for the one-timbre and three-timbre conditions for melodies and sentences. Data represent the grand average across time based on each participant's hemisphere with the largest 50 ms average around the peak in the time window of interest for (A) Melodies (ERAN: 150–250 ms), and (B) Sentences (LAN: 270–360 ms). Note that this data does not represent the hemisphere with the largest response in the ELAN time window. Shaded error bars indicate one standard error either side of the mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
ERP Mean Amplitudes and Standard Deviations for the 50 ms Average around the Peak in the Time Window of Interest (indicated in brackets).

	One-timbre		Three-timbres	
Stimuli	*M*	*SD*	*M*	*SD*
Melodies (150–250 ms)	− 3.41	2.40	− 1.98	1.54
Sentences (100–150 ms)	− 1.39	2.32	− 0.62	1.58
Sentences (270–360 ms)	− 2.91	2.94	− 2.16	2.87

time window was visually reduced in the three-timbre condition; however, there was no significant difference between the one-timbre condition and the three-timbre condition, $t(21) = 1.04$, $p = .31$. The lack of significance appears to be due to large variance within our sample, which will be discussed further in the discussion.

2.4. Hemisphere analysis

2.4.1. Melodies

For the brain response to melodies in the ERAN time window, there was a significant main effect of timbre, $F(21) = 6.50$, $p = .02$, $\eta^2 = 0.24$, a significant main effect of hemisphere, $F(21) = 27.5$, $p < .001$, $\eta^2 = 0.57$, and no interaction between timbre and hemisphere, $F(21) = 0.06$, $p = .81$. Further analyses revealed that the brain response in the right hemisphere (RH) was significantly more negative than the brain response in the left hemisphere (LH) in both the one-timbre condition (right: $M = -3.26$, $SD = 2.43$; left: $M = -2.39$, $SD = 2.38$), $t(21) = 3.45$, $p' = .01$, $d = 0.74$, and the three-timbre condition (right: $M = -1.87$, $SD = 1.54$; left: $M = -1.08$, $SD = 1.46$), $t(21) = 3.63$, $p' = .01$, $d = 0.78$. Further, the one-timbre condition was significantly more negative than the three-timbre condition in both the RH, $t(21) = 2.56$, $p' = .04$, $d = 0.57$, and the LH, $t(21) = 2.30$, $p' = .04$, $d = 0.51$. These results show that the ERAN was stronger in the RH, as suggested by previous research (Koelsch, 2013).

2.4.2. Sentences

For the brain response to sentences in the ELAN time window, there was no main effect of timbre, $F(21) = 1.71$, $p = .21$, a main effect of hemisphere, $F(21) = 4.9$, $p = .04$, $\eta^2 = 0.19$, and no interaction between timbre and hemisphere, $F(21) = 0.27$, $p = .61$. Further analyses showed no differences between the one-timbre condition in the RH ($M = -1.13$, $SD = 2.29$) compared to the LH ($M = -0.83$, $SD = 2.41$), $t(21) = 1.35$, $p' = .53$, and no differences in the three-timbre condition in the RH ($M = -0.41$, $SD = 1.70$) compared to the LH ($M = 0.03$, $SD = 1.63$), $t(21) = 2.06$, $p' = .21$. There were also no significant differences between the one- and three-timbre conditions in the RH, $t(21) = 1.15$, $p' = .53$, or the LH, $t(21) = 1.39$, $p' = .53$. This analysis suggests that overall, the brain response was slightly lateralized to the RH; however, this effect was quite weak and was not evident between conditions.

For the brain response to sentences in the LAN time window, there was no main effect of timbre, $F(21) = 1.20$, $p = .29$, no main effect of hemisphere, $F(21) = 3.11$, $p = .09$, and no interaction between timbre and hemisphere, $F(21) = 0.01$, $p = .91$.

3. Discussion

The current ERP experiment investigated whether behavioral and electrophysiological responses to syntactic violations in music and language were reduced when syntactic sequences were disrupted with alternating timbres (three-timbre condition) compared to when they were within one auditory stream (one-timbre condition). For melodies, behavioral data showed that participants were significantly more sensitive to syntactic violations in the one-timbre condition compared to the three-timbre condition. This finding was also reflected in the ERP results, with the ERAN response to out-of-key notes significantly reduced when melodies were played with three alternating instruments compared to only one instrument. This finding suggests that alternating timbres affect the processing of music syntax in the brain, likely due to an interruption of auditory streaming processes at an early stage of processing. For spoken sentences, we did not observe a significant behavioral or electrophysiological difference between the one- and three-timbre conditions, although the left anterior negativity (LAN) ERP response was attenuated in the three-timbre condition compared to the one-timbre condition.

3.1. Music syntax and timbre

Previous behavioral research has suggested that alternating timbres in a musical sequence reduces processing of syntactic structure (Bregman, 1990; Fiveash et al., under review; McAdams, 1999). However, the current investigation is the first to investigate this phenomenon with ERPs, which allowed us to investigate the effects of timbre on the neural processing of syntax in real time. The ability to detect syntactic violations requires the brain to continuously track incoming information, and to register when there is an element that does not adhere to the tonal context. Despite the apparent sophistication of this process, the operation occurs automatically and without overt attention to the stimuli (Loui et al., 2005). In the current experiment, participants exhibited the ERAN in response to out-of-key notes in both the one- and three-timbre conditions, suggesting that the out-of-key note was registered in both conditions. At the group-level, the ERAN was right lateralized in both the one- and three-timbre conditions. What is interesting is that the ERAN response was significantly reduced when the melodies were played by three timbres compared to one timbre, showing a direct influence of timbre on syntactic processing. The reduced brain response to syntactic violations in the three-timbre condition helps to explain our behavioral results, which suggest that participants were less sensitive to out-of-key notes in the three-timbre condition than in the one-timbre condition.

The reduced brain response to a syntactic violation when the melody is played with three timbres may be due to the disrupting effect of timbre changes on auditory streaming. This in turn affects syntactic structure building, leading to a less coherent melody and a weaker syntactic representation. Perceptual streaming accounts (Bregman, 1990; Cusack and Roberts, 2000; Iverson, 1995) and Gestalt principles (Deutsch, 2013) suggest that incoming auditory streams are grouped together by similarity (e.g., timbre) and proximity (e.g., pitch distance). Furthermore, models of music perception and auditory sentence processing show an initial feature extraction and acoustic analysis stage where timbral information is processed (Friederici, 2002; Koelsch, 2011). By disrupting a salient similarity cue (timbre) in early stages of perceptual analysis, and placing a larger burden on auditory streaming processes, it is likely our stimuli made it more difficult for participants to group notes into a coherent stream. However, grouping was not prevented entirely, as participants were able to detect violations in both conditions. It is possible that other grouping principles, such as pitch proximity and regular timing, promoted partial streaming of the melodic information. We therefore suggest that the strength of syntactic representations in the brain is directly related to early auditory streaming processes.

With alternating timbres rendering the melody less coherent, it is possible that predictive processes were also less efficient. Prediction is an important element in both music and language, and can operate on multiple levels (Patel and Morgan, 2016). An out-of-key note in a one-timbre sequence is more unexpected than an out-of-key note in a three-timbre sequence, as the rest of the stream is expected and easily predicted. In a three-timbre context, the timbre of the melodic stream is less predictable, and so the brain may hold weaker predictions about upcoming elements in relation to syntax as well. When these predictions are violated, it may come as less of a surprise. Overall, the current experiment shows that alternating timbres disrupt the brain's ability to fully process syntactic errors in music, at the level of both behavior and the brain. It may be valuable in future research to investigate the effects of other methods of disrupting auditory streaming. If the current ERP results reflect a disruption to processes of auditory streaming, then any manipulation that disrupts auditory streaming should lead to a similar reduction in brain response to syntactic violations.

Alternatively, it is possible that changes in timbre resulted in shifts in attention that distracted participants from the out-of-key notes or drew attentional resources away from auditory streaming processes (Jones et al., 2010). The connection between attention and auditory streaming is complex and not well understood (Sussman et al., 2007). Auditory streaming is largely considered a bottom-up process, and timbre a bottom-up cue to stream segregation (Bregman, 1990; Disbergen et al., 2018). However, it has been shown that auditory streaming can be affected by attention and top-down processes (see Cusack et al., 2004). Therefore, it is possible that attention affected the formation of auditory streams in the current experiment.

We suggest four reasons why this explanation is unlikely in the current experiment. First, participants were told to focus on the melodic content as opposed to the alternating timbres, and their task was related to the melody and not the timbres. Therefore, it can be expected that top-down attention was directed to the melodies as opposed to the timbre changes. Second, an extensive body of research suggests that the processing of timbre and the formation of auditory streams are inherently linked (Bregman, 1990). Therefore, changes in timbre have a direct impact on the formation and coherence of auditory streams, independent of attention (see *primitive auditory streaming* in Bregman, 1990). Third, syntactic processing occurs automatically, even when participants are not paying attention to the stimulus (though attention does impact this process, Loui et al., 2005). Thus, even if participants were distracted by timbre changes, if the incoming sequences were perceived as coherent streams, then participants should still have had a strong response to the out-of-key note. Fourth, our previous research revealed that when melodies and sentences were presented concurrently, melodies with alternating timbres led to *reduced* interference by those melodies on recall of accompanying sentences. If alternating timbres were generally distracting (leading to shifts in attention), then we would have expected *greater* interference by melodies on recall of accompanying sentences (Fiveash et al., under review). Future research should continue to investigate the links between auditory streaming, attention, and timbre, as an important insight into perceptual processing.

3.2. Language syntax and timbre

The similarities between music and language in relation to syntax led to the prediction that three timbres in language (three voices) may also reduce the brain's response to syntactic violations compared to one timbre (one voice). We predicted that we would observe the ELAN in response to phrase-structure violations, as seen in previous literature (Friederici, 2002; Maidhof and Koelsch, 2011). This prediction was only partially supported, since we found a small but statistically reliable ELAN to syntactic violations in sentences in the one-timbre condition but not the three-timbre condition. In contrast, the LAN offered a more reliable response—with a larger peak within the 270–360 ms time window in both the one-timbre and three-timbre conditions. Our analysis showed that the LAN in the three-timbre condition was reduced compared to the one-timbre condition, though this difference was not significant. This lack of significance appears to be due to the large amount of variation between participants.

Finding a statistically reliable effect of timbre changes for music stimuli, but not language stimuli, was not predicted. There are at least six potential explanations for this unexpected finding. First, repeated exposure to conventional Western instruments may have led to expectations for a high level of consistency in timbre for different events within a musical stream, such that changes in timbre readily disrupt processes of auditory streaming. In contrast, we may be more tolerant to changes in vocal timbre within a given speech stream, because speakers routinely use such changes in vocal timbre as part of prosodic communication. More generally, auditory sentence processing is inherently variable, as we have to process words, prosody, and semantics in addition to syntax, which could have led to a noisier signal.

Second, in the current experiment, the music stimuli were isochronous and in 4/4 timing, and hence highly predictable. Our language stimuli, in contrast, may have sounded rhythmically unnatural, thereby obscuring the effect.

Third, it is possible that timbre is more important to syntactic processing for music than for language, as cues to timbre are not as indicative to meaning in language as they are to music. Meaning in language is delivered irrespective of timbre, due to the referential and propositional nature of language (Jackendoff, 2009). Meaning in music on the other hand is a complex phenomenon, related to a number of aspects of the music, including pitch and timbre (Koelsch et al., 2004). Because of this distinction in the way meaning is communicated in music and language, timbral cues may contribute less to the processing of syntax in language than in music.

A fourth consideration is that the grammatical errors were more obvious in the language stimuli, as evidenced by ceiling effects in our behavioral data. It is possible that we did not see a difference between the one- and three-timbre conditions because the task was too easy in comparison with the music task.

A fifth possible reason why we did not find any effect in the language condition could be due to our stimuli. The stimuli were created to ensure that there was a baseline of silence before the critical word so that the ERP to the critical word was not affected by the previous word. This manipulation may have led to unnatural sounding speech that could have resulted in "noisy" brain activity.

Sixth, the speakers were all female with Australian accents. It is possible that spectral differences between speaker voices were not as large or noticeable as spectral differences between the three musical timbres. The smaller variation in speaker voices may therefore account for the non-significant difference in the LAN between the one- and three-timbre conditions, as the variation may not have been large enough to implicate separate sources in auditory streaming. However, it should be emphasized that the three voices were clearly discriminable, and unlikely to have been perceived by any of our participants as arising from the same speaker. Therefore, it is unlikely that the lack of significant findings can be explained by participants processing the three distinct voices as though they arose from the same source.

To continue to investigate links between language syntax and timbre, future research should aim for a larger signal to noise ratio by increasing the number of trials. In addition, it may be valuable to explore different methods of disrupting auditory streams in language stimuli, experiment with different voice timbres, and introduce more sensitive grammatical errors to try and observe the effect. For example, obvious syntactic errors may be easily perceived regardless of timbre. It would also be interesting to see if the timbre effect occurs in relation to semantic errors in language. If the current study were to be repeated, it would also be possible to design sentences where every word ends on a "stop" consonant (e.g., k, t, p), so that words do not run together. This manipulation would make it easier to splice different voices together without a pause between words. Further, by increasing spectral differences between speakers, or manipulating speech artificially, it might also be possible to induce greater disruptions to auditory streaming, resulting in a larger difference between the one- and three-timbre conditions.

4. Conclusion

The current experiment shows, for the first time, that the brain response to syntactic violations in music is reduced when melodies are played by three timbres compared to one timbre. Within a music perception framework, this finding suggests that alternating timbres disrupt auditory streaming processes in an initial feature extraction stage, which in turn leads to impaired syntactic structure building processes. Although the same pattern was observed in sentence processing, the difference was not significant, likely due to high individual variation in brain responses to auditory sentences. It would be useful if future studies could further explore brain responses to syntactic violations in speech by using carefully controlled stimuli, and increasing the signal to noise ratio. It would also be useful to see whether the current findings for the music stimuli can be generalised to different timbres.

Acknowledgements

This research was supported by the Macquarie University Research Excellence Scholarship (MQRES), awarded to the first author (allocation number 2013087), and an Australian Research Council (ARC) Discovery Grant (DP130101084) awarded to the second author.

References

Boersma, P., 2001. Praat, a system for doing phonetics by computer. Glot Int. 5, 341–347.

Boucher, R., Bryden, M.P., 1997. Laterality effects in the processing of melody and timbre. Neuropsychologia 35 (11), 1467–1473. http://dx.doi.org/10.1016/S0028-3932(97)00066-3.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10 (4), 433–436.

Bregman, A.S., 1990. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press, Cambridge, MA.

Coulson, S., King, J.W., Kutas, M., 1998. Expect the unexpected: event-related brain response to morphosyntactic violations. Lang. Cogn. Process. 13 (1), 21–58. http://dx.doi.org/10.1080/016909698386582.

Cusack, R., Roberts, B., 2000. Effects of differences in timbre on sequential grouping. Percept. Psychophys. 62 (5), 1112–1120. http://dx.doi.org/10.3758/BF03212092.

Cusack, R., Deeks, J., Aikman, G., Carlyon, R.P., 2004. Effects of location, frequency region, and time course of selective attention on auditory scene analysis. J. Exp. Psychol. Hum. Percept. Perform. 30 (4), 643–656. http://dx.doi.org/10.1037/0096-1523.30.4.643.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

Deutsch, D., 2013. The Psychology of Music, 3rd ed. Academic Press, San Diego.

Disbergen, N.R., Valente, G., Formisano, E., Zatorre, R.J., 2018. Assessing top-down and bottom-up contributions to auditory stream segregation and integration with polyphonic music. Front. Neurosci. 121. http://dx.doi.org/10.3389/fnins.2018.00121.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., Gibson, E., 2009. Structural integration in language and music: evidence for a shared system. Mem. Cogn. 37 (1), 1–9. http://dx.doi.org/10.3758/MC.37.1.1.

Fiveash, A., Pammer, K., 2014. Music and language: do they draw on similar syntactic working memory resources? Psychol. Music 42 (2), 190–209. http://dx.doi.org/10.1177/0305735612463949.

Fiveash, A., McArthur, G., & Thompson, W. F. (2018). Language processing in the presence of music: An investigation of interference effects (under review).

Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. Trends Cogn. Sci. 6 (2), 78–84. http://dx.doi.org/10.1016/S1364-6613(00)01839-8.

Garza Villarreal, E.A., Brattico, E., Leino, S., Østergaard, L., Vuust, P., 2011. Distinct neural responses to chord violations: a multiple source analysis study. Brain Res. 1389, 103–114. http://dx.doi.org/10.1016/j.brainres.2011.02.089.

Gunter, T., Friederici, A.D., Schriefers, H., 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. J. Cogn. Neurosci. 12 (4), 556–568.

Hagoort, P., Wassenaar, M., Brown, C.M., 2003. Syntax-related ERP-effects in Dutch. Cogn. Brain Res. 16 (1), 38–50. http://dx.doi.org/10.1016/S0926-6410(02)00208-2.

Huron, D., 2008. Sweet Anticipation: Music and the Psychology of Expectation. MIT Press, Cambridge, MA.

Iverson, P., 1995. Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. J. Exp. Psychol. Hum. Percept. Perform. 21 (4), 751–763. http://dx.doi.org/10.1037/0096-1523.21.4.751.

Jackendoff, R., 2009. Parallels and nonparallels between language and music. Music. Percept. 26 (3), 195–204. http://dx.doi.org/10.1525/mp.2009.26.3.195.

Jentschke, S., Koelsch, S., Friederici, A.D., 2005. Investigating the relationship of music and language in children: influences of musical training and language impairment. Ann. N. Y. Acad. Sci. 1060, 231–242. http://dx.doi.org/10.1196/annals.1360.016.

Jones, D.M., Hughes, R.W., Macken, W.J., 2010. Auditory distraction and serial memory: the avoidable and the ineluctable. Noise Health 12 (49), 201–209. http://dx.doi.org/10.4103/1463-1741.70497.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., 2007. What's new in psychtoolbox-3. Perception 14, 1–16. Retrieved from. https://nyuscholars.nyu.edu/en/publications/whats-new-in-psychtoolbox-3.

Koelsch, S., 2011. Towards a neural basis of music perception - a review and updated model. Front. Psychol. 110. http://dx.doi.org/10.3389/fpsyg.2011.00110.

Koelsch, S., 2013. Brain and Music. John Wiley & Sons, Oxford, UK.

Koelsch, S., Jentschke, S., 2008. Short-term effects of processing musical syntax: an ERP study. Brain Res. 1212, 55–62. http://dx.doi.org/10.1016/j.brainres.2007.10.078.

Koelsch, S., Gunter, T., Friederici, A.D., 2000. Brain indices of music processing: "nonmusicians" are musical. J. Cogn. Neurosci. 12 (3), 520–541.

Koelsch, S., Gunter, T., Schroger, E., Tervaniemi, M., Sammler, D., Friederici, A.D., 2001. Differentiating ERAN and MMN: an ERP study. Neuroreport 12 (7), 1385–1389.

Koelsch, S., Gunter, T., v. Cramon, D., Zysset, S., Lohmann, G., Friederici, A.D., 2002. Bach speaks: A cortical "language-network" serves the processing of music. NeuroImage 17 (2), 956–966. http://dx.doi.org/10.1006/nimg.2002.1154.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., Friederici, A.D., 2004. Music, language and meaning: brain signatures of semantic processing. Nat. Neurosci. 7 (3), 302–307. http://dx.doi.org/10.1038/nn1197.

Koelsch, S., Gunter, T., Wittfoth, M., Sammler, D., 2005. Interaction between syntax processing in language and music: an ERP study. J. Cogn. Neurosci. 17 (10), 1565–1577.

Koelsch, S., Heinke, W., Sammler, D., Olthoff, D., 2006. Auditory processing during deep propofol sedation and recovery from unconsciousness. Clin. Neurophysiol. 117 (8), 1746–1759. http://dx.doi.org/10.1016/j.clinph.2006.05.009.

Kunert, R., Willems, R.M., Casasanto, D., Patel, A.D., Hagoort, P., 2015. Music and language syntax interact in Broca's area: an fMRI study. PLoS One 10 (11), e0141069. http://dx.doi.org/10.1371/journal.pone.0141069.

Levitin, D.J., Menon, V., 2003. Musical structure procesed in "language" areas of the brain: a possible role for Brodmann area 47 in temporal coherence. NeuroImage 20, 2142–2152. http://dx.doi.org/10.1016/S1053-8119(03)00482-8.

Loui, P., Grent-'t-Jong, T., Torpey, D., Woldorff, M., 2005. Effects of attention on the neural processing of harmonic syntax in Western music. Cogn. Brain Res. 25 (3), 678–687. http://dx.doi.org/10.1016/j.cogbrainres.2005.08.019.

Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique, 2nd ed. MIT Press, USA.

Maess, B., Koelsch, S., Gunter, T., Friederici, A.D., 2001. Musical syntax is processed in Broca's area: an MEG study. Nat. Neurosci. 4 (5), 540–545.

Mahajan, Y., McArthur, G., 2010. Does combing the scalp reduce scalp electrode impedances? J. Neurosci. Methods 188 (2), 287–289. http://dx.doi.org/10.1016/j.jneumeth.2010.02.024.

Maidhof, C., Koelsch, S., 2011. Effects of selective attention on syntax processing in music and language. J. Cogn. Neurosci. 23 (9), 2252–2267.

Masataka, N., 2009. The origins of language and the evolution of music: a comparative perspective. Phys Life Rev 6 (1), 11–22. http://dx.doi.org/10.1016/j.plrev.2008.08.003.

McAdams, S., 1999. Perspectives on the contribution of timbre to musical structure. Comput. Music. J. 23 (3), 85–102. http://dx.doi.org/10.1162/014892699559797.

McAdams, S., 2013. Musical timbre perception. In: Deutsch, D. (Ed.), Psychology of Music. Elsevier, Inc., USA.

Miranda, R.A., Ullman, M.T., 2007. Double dissociation between rules and memory in music: an event-related potential study. NeuroImage 38 (2), 331–345. http://dx.doi.org/10.1016/j.neuroimage.2007.07.034.

Neville, H., Nicol, J.L., Barss, A., Forster, K.I., Garrett, M.F., 1991. Syntactically based sentence processing classes: evidence from event-related brain potentials. J. Cogn. Neurosci. 3 (2), 151–165. http://dx.doi.org/10.1162/jocn.1991.3.2.151.

Ono, K., Nakamura, A., Yoshiyama, K., Kinkori, T., Bundo, M., Kato, T., Ito, K., 2011. The effect of musical experience on hemispheric lateralization in musical feature processing. Neurosci. Lett. 496 (2), 141–145. http://dx.doi.org/10.1016/j.neulet.2011.04.002.

Patel, A.D., 2003. Language, music, syntax and the brain. Nat. Neurosci. 6 (7), 674–681.

Patel, A.D., 2008. Music, Language, and the Brain. Oxford University Press, New York.

Patel, A.D., Morgan, E., 2016. Exploring cognitive relations between prediction in language and music. Cogn. Sci. 303–320. http://dx.doi.org/10.1111/cogs.12411.

Patel, A.D., Gibson, E., Ratner, J., Besson, M., Holcomb, P., 1998. Processing syntactic relations in language and music: an event-related potential study. J. Cogn. Neurosci. 10 (6), 717–733.

Reiterer, S., Erb, M., Grodd, W., Wildgruber, D., 2008. Cerebral processing of timbre and loudness: fMRI evidence for a contribution of Broca's area to basic auditory discrimination. Brain Imaging Behav. 2 (1), 1–10. http://dx.doi.org/10.1007/s11682-007-9010-3.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H.J., ... Schulze-Bonhage, A., 2013. Co-localizing linguistic and musical syntax with intracranial EEG. NeuroImage 64, 134–146. http://dx.doi.org/10.1016/j.neuroimage.2012.09.035.

Stanislaw, H., Todorov, N., 1999. Calculation of signal detection theory measures. Behav. Res. Methods Instrum. Comput. 31 (1), 137–149.

Steinbeis, N., Koelsch, S., 2008. Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. Cereb. Cortex 18 (5), 1169–1178. http://dx.doi.org/10.1093/cercor/bhm149.

Sussman, E.S., Horváth, J., Winkler, I., Orr, M., 2007. The role of attention in the formation of auditory streams. Percept. Psychophys. 69 (1), 136–152. http://dx.doi.org/10.3758/BF03194460.