

Part VIII

Free Talk Session 5

EFFECT OF SOUND ON MEMORY AND IMPRESSIVENESS OF VISUAL IMAGERY

Natalia Postnova and Shin-ichiro Iwamiya

*Communication Design Science Course, Department of Design
Graduate School of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku
Fukuoka 815-8540, Japan*

<nataliapostnova@gmail.com, iwamiya@design.kyushu-u.ac.jp>

Abstract

To clarify the functions of sounds in multimedia and the effect of sound on memory of visual imagery, the psychophysical experiments were conducted. The experimental stimuli were a series of symbolic pictures presented in video sequences with or without sound, or half with/half without sound. After every sequence participants were asked to recall the symbols they saw by selecting the right ones from the list. Experiments showed that the number of correct answers was larger when all the pictures in the sequence were presented with sound than when all or half of the pictures in the sequence were presented without sound. The same tendency was noticed when half of the symbols in the sequence were presented with sound and half without, the participants reported a higher number of symbols that were presented with sound. Furthermore, in the same condition, the number of selected pictures as the most impressive was larger when the pictures were presented with sound. These results suggested that sounds in multimedia have functions to emphasize the visual impression and facilitate visual memory.

The development of the contemporary media raised a lot of questions on the effect of sound on perceived imagery information. The memorability and impressiveness of the visual information is an important aspect of media-production. Can sound boost our memory or on the contrary distract us from the visual information? The increasing use of multimedia materials for various purposes including educational, therapeutic, recreational and others makes these questions even more important to answer.

Previous research in this area showed a complex relation between sound as a distracter and sound as a supporter of visual information. On one hand, we know that speech and non-speech stimuli of non-steady state can disrupt the performance of simple cognitive tasks (Jones et al., 1993), but on the other hand congruent sound can benefit the perception of visual information (Vroomen et al., 2000). However, the effect of the congruent, but non-steady sound is questionable. The purpose of the current study is to clarify the effect of a simple sound on memory and impression of visual imagery. To measure the expected effect, a simple memory span test is used (Miller et al., 1956; Ellermeier et al., 2015). We expect to detect difference in performance depending on the presentation of the visual imagery—with sound or without sound. As the presentation of the visual information in the memory span tests is usually periodical—the information appears on the screen one by one over equal periods of time, the sound synchronized with every element is steady periodical sound. To distinguish between the effect of the steady and non-steady sound, the third condition is used. In that condition, not all of the visual elements of memory span test are presented with the sound, but just a half of them. That way we can distinguish the effect of the sound itself, and the effect of steady and non-steady sounds on the performance of the cognitive task.



Fig. 1. Symbolic images used in the experiments.

Selection and designing of materials for visual stimuli

In order to conduct the experiments, the visual and sound materials for the memory span test had to be designed. As the test was based on the standard memory span test, the visual information for memorization should have been decided. Usually the digits are used—they are presented in a sequence of 3 to 10 digits once (depends on participant aptitude), and participants are asked to recall it in a correct order. During the pre-experimental stage, several stimuli were tested—digits, roman letters (consonants only) and black-and-white symbolic images. In the cases of digits and roman letters, participants were found to use various memorization techniques—for example, vocalization of the digits or letters and making meaning of the sequences. That consequently compromised the results. In the case of symbolic images, uses of such techniques were not found, hence the black-and-white symbolic images were chosen as visual stimuli for the experiments. The images are shown in Figure 1 (Fig. 1).

In the final design, 20 symbolic images were used to create 12 random sequences of 8 symbols each. For each sequence 3 sound conditions were created:

1. No sound (NS condition): control condition.
2. Fully synchronized sound condition (SS condition): Every symbol appearance was synchronized with the sound—440-Hz pure tone (0.1 second).
3. Half synchronized sound condition (HSS condition): Randomly chosen half of the symbols in the sequence were presented with sound (the same as in SS condition), while others without.

Figure 2 (Fig. 2) shows the schematic representation of SS and HSS conditions.

All together there were 36 videos—12 sequences of 8 symbols each in 3 different conditions. The same video material was used for 1st and 2nd experiments. Both experiments were conducted in soundproof rooms, and the sound stimuli were presented using headphones with the A-weighted sound pressure level of 67dB in both ears.

The experiments consisted of 3 sessions with 5 minutes breaks in between. Every session consisted of 12 different videos in different conditions, so during 3 sessions all 12 sequences were presented in all 3 conditions.

Experiment 1: The effect of sound on memorization of visual imagery

The purpose of the first experiment was to detect the effect of sound on short term memory. The participants were asked to watch the sequences and try to memorize the

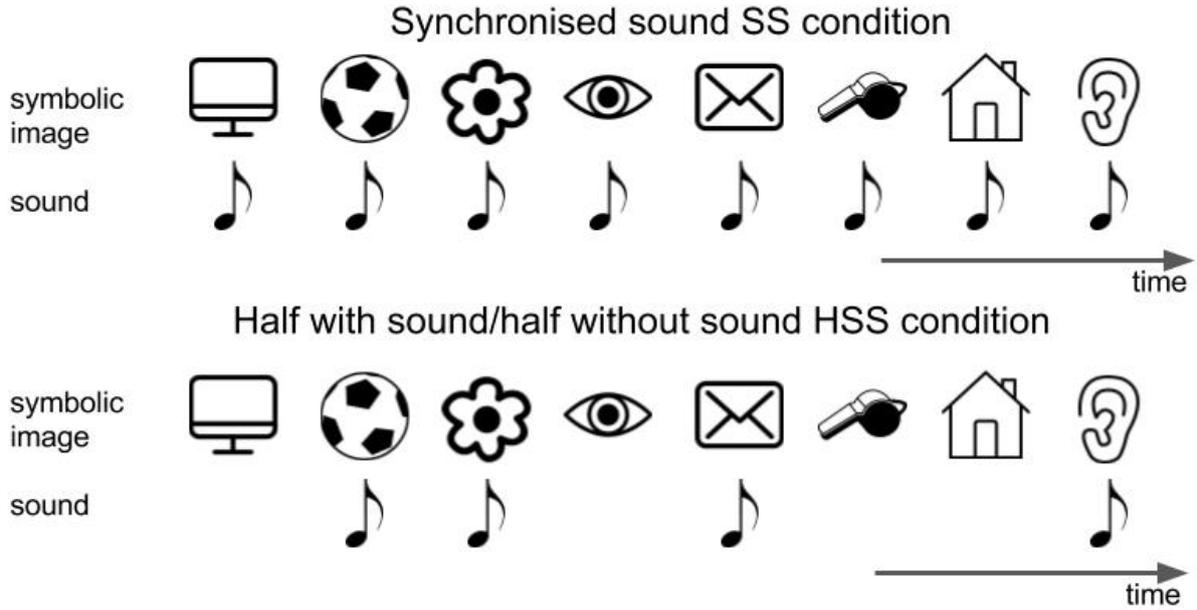


Fig. 2. Sound conditions in SS and HSS.

symbols (order was not important). After each sequence, they were asked to mark the symbols they remembered on the answer blank.

The scores were calculated based on the number of correct symbols recalled, the results were grouped based on the sound condition the sequence was presented in. To compare performances in different conditions, the absolute score for each participant was normalized against the controlled condition—NS (no sound) condition. For each participant, the result of performance in NS conditions in the session was calculated and performances in other conditions—SS and HSS, were measured against its score in %. This normalization was used in order to eliminate individual differences of performance. The average performance among all the participants in different groups using normalized scores are presented in a table (Table 1).

Because the 95% confidence interval of average value in SS condition does not include 100, the difference of performance between SS and NS condition is statistically significant. As it can be seen from table (Table 1), participants performed higher in SS condition. On the other hand, HSS condition did not have the same effect. The performance level in HSS is equivalent to that in NS condition. The difference of average

Table 1. The average performance in different conditions in percentages to NS condition (HSS—half synchronized sound condition, SS—fully synchronized sound condition, NS—no sound controlled condition)

Condition	Average ($\pm 95\%$ confidence interval)
HSS	99 (± 4)
SS	108 (± 7)
NS	100

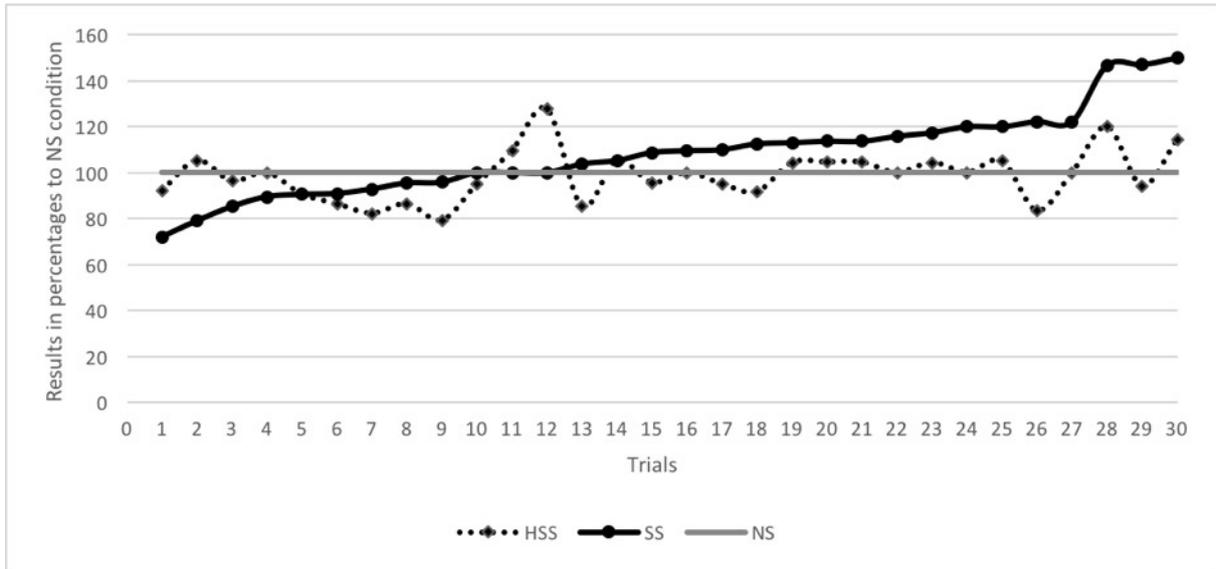


Fig. 3. Memory test performances results in each trial. HSS—half-synchronized condition, SS—fully synchronized condition.

values between SS and HSS condition is statistically significant, $t(47) = -2.4, p = 0.02$.

The comparison graph, showing the result of every trial in SS and HSS conditions compared with NS conditions is presented in Figure 3. As it can be seen from the graph, the SS condition on average has better results than performances in other conditions—in 23 out of 30 trials performance in SS was better than performance in HSS, and in 18 out of 30 trials performance in SS was better than in controlled NS condition.

In HSS condition the effect of sound on memory can be compared within the same trials, as half of the stimuli were presented with sound and half without. These two numbers—the number of recalled symbols that were presented with sound and the number of recalled symbols that were presented without the sound, were calculated separately. This analysis showed that the percentage of correctly recalled symbols with sound was higher than the percentage of symbols without sound. These numbers are presented in Figure 4.

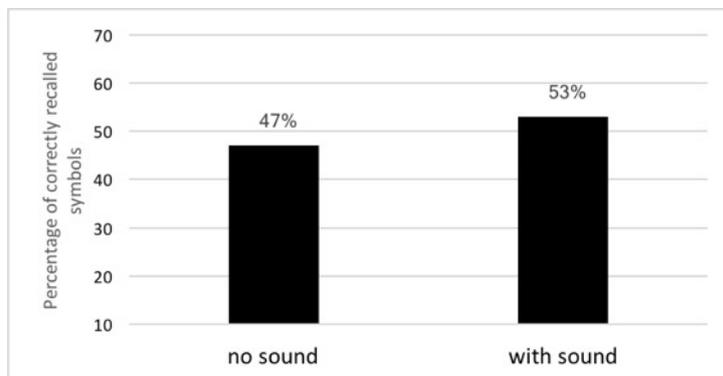


Fig. 4. Correctly recalled symbols with sound and without sound in HSS condition (% from overall number of the recalled symbols).

It is seen that participants reported on average 6% more symbols presented with sound than those presented without the sound, and the difference between them is statistically significant, $t(70) = 2.6$, $P = 0.011$.

Based on these results, we can conclude that sound affected the performance of the task of the memorization of visual stimuli. Results showed that sound in general made the visual stimuli more memorable. The steady fully synchronized congruent sounds boosted the performance of the task, while non-steady sound did not have such effect. However, the tendency to memorize stimuli with the sound better remained even in case of non-steady sounds.

Experiment 2: Effect of sound on the impressiveness of visual imagery

The purpose of the 2nd experiment was to clarify the effect of a sound on the impressiveness of the visual imagery. For this experiment, the same visual materials as in experiment 1 were used. The participants were asked to watch the video sequences without trying to memorize the symbols, and after watching draw 3 symbols that seemed to be most impressive of all of them.

The results were analyzed with the following method. Symbols in every video sequence were separated in 2 groups based on how particular symbols in particular sequences were presented in HSS (half synchronized) condition: S group (with sound) and N group (without sound). The presentation condition of symbols of S and N groups was different only in HSS condition; in SS condition both groups were presented with the sound and in NS condition both groups were presented without sound. So that way if there is no difference between the number of drawn symbols from S-group and N-group in conditions NS and SS, but there is in condition HSS, then we can say that it is the result of the effect of the sound.

After the experiments, all the data was analyzed using the following method. For every participant the number of drawn symbols from S-group and the number of symbols from N-group was calculated for every condition and converted into percentages. The average percentages for each group among the participants are presented in Figure 5.

The graph shows that among the drawn symbols in HSS condition there were 20% more symbols that were presented with sound (drawn symbols from S group) than the ones that were presented without sound (drawn symbols from N group), and this difference was statistically significant, $t(11) = 4.3$, $P < 0.01$. As the same sequences were presented in different sound conditions it would be logical to assume that if the symbols themselves were more impressive because of some other features rather than the sound, we would see the same tendency in other conditions. However, the graph shows that in condition SS and NS the proportion of the group of symbols that were presented with or without sound in condition SN are almost 50/50, and the differences between S and N groups were not statistically significant. Therefore, we can conclude that synchronized sound affect the impressiveness of the visual imagery.

Discussion

Altogether 30 trials were conducted for each of the experiments, and performances were analyzed using statistical analysis. The results showed that fully synchronized congruent sounds boosted the performance of simple memory task. In the follow up interview after the experiment, some participants reported that NS (no sound) condition was the most

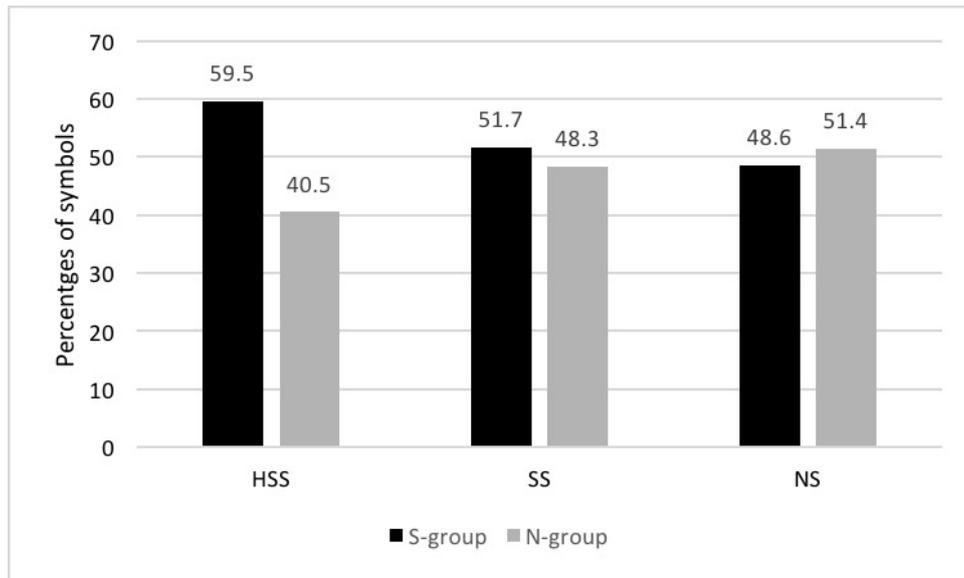


Fig. 5. The percentage of drawn symbols from S (with sound) or N (without sound) groups in three different conditions. HSS—half synchronized sound, SS—fully synchronized sound, NS—no sound.

complicated for them as it was difficult to concentrate, while others reported that HSS (partially synchronized) condition caused the most confusion and distraction, as non-steady sound was “annoying.” No one reported SS condition as the most difficult one, though some participants mentioned that they preferred the condition without any sound.

Based on the results, it was suggested that we dealt with two different effects here. One is the effect of sound itself and another is the effect of periodicity of sound. It is known that non-periodic sounds are much more distractive than periodic ones, and non-periodic sounds attract our attention much more than periodic ones to the sound itself (Jones et al., 1993). We could see that effect when comparing half synchronized HSS condition and fully synchronized SS condition—participants performed worse in HSS than SS condition. It seems that periodic sounds boosted steady attention to the visual elements synchronized with it, because the performance in SS condition was higher than in no sound NS condition. Non-steady sounds in condition HSS also had the similar effect—we could see higher number of the symbols that were presented with sound reported, but as it was said earlier, in general non-periodic sound had more of a distracting effect and the performance was not improved by half synchronized sound compare to the NS condition. The further research can be conducted to confirm or reject these assumptions.

The second experiment showed the effect of sound on the impressiveness of symbolic images. The impression was stronger when images were presented with sound. This might facilitate the function of short term memory. The results of the 1st experiment support that: the number of recalled symbols presented with sound in HSS condition was larger than that without sound; the performance difference between SS and NS condition also showed the function of sound to facilitate the function of short term memory.

Another direction of further research could be focused on differences among participants, as some participants showed much higher influence to different sound conditions than others. It seems like some people are much more sensitive to sound, while others showed very small difference in performance. Further research can be done in order to

clarify whether certain personal qualities, past experience or skills affect person's sensitivity to it.

References

- Ellermeier, W., Kattner, F., Ueda, K., Doumoto, K. & Nakajima, Y. (2015). Memory disruption by irrelevant noise-vocoded speech: Effects of native language and the number of frequency bands. *The Journal of the Acoustical Society of America*, 138, 1561–1569.
- Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory and Cognition*, 21(3), 318–366.
- Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5), 1583–1590.

INVESTIGATING DIFFERENCES IN THE DYNAMIC-SYSTEMS STRUCTURE OF AUDITORY COGNITION AS A FUNCTION OF MUSIC TRAINING

Naomi du Bois¹, José M. Sanchez Bornot², KongFatt Wong-Lin², Mark A. Elliott^{1*}, and
Girijesh Prasad²

¹*School of Psychology, National University of Ireland, Galway, Ireland*

²*Intelligent Systems Research Centre, University of Ulster, Magee Campus, Derry, NI*

*Correspondence: <mark.elliott@nuigalway.ie>

Abstract

A magnetoencephalographic (MEG) investigation of differences in the structure of the dynamic auditory cognition system, dependent on music training as opposed to no experience playing an instrument, was conducted using an auditory priming paradigm designed by Aksentijevic, Barber and Elliott [(2011). JEP: Human Percept. Perform., 37, 1628]. This paradigm employed stimulus entrainment to evoke an auditory gamma-band response (aGBR, i.e. an oscillatory response in the range 30-70 Hz) that is phase locked to the stimulus. Frequencies in this range have been demonstrated to facilitate a response to a deviant stimulus (inharmonic) depending on their relationship in phase with a slower theta rhythm [Elliott (2014). Front. Psychol., 5, 990]. Neuroscientific research has demonstrated that syntactically irregular chords elicit event related potentials (ERPs) with negative polarity and peak latencies of around 150-350 ms post stimulus onset [Rohrmeier & Koelsch (2012). Int. J. Psychophysiol., 83, 165-175]. The focus of the time frequency analyses for this was on the effect of priming on these auditory responses as a function of musical experience.

Auditory binding refers to the integration of sounds to form a whole sound object from different acoustical components that are coded in anatomically separate parts of auditory cortex. This may be achieved through the synchronized activity of neural assemblies across the separate cortical loci with an emergent oscillatory code serving to bind neurons into a single dynamic assembly. Our research used an auditory priming paradigm developed by Aksentijevic, Barber, and Elliott (2011) to examine potential differences in the oscillatory codes involved in this process of binding, dependent on musical training. The paradigm primes the auditory system with sound stimuli designed to evoke an auditory gamma-band response (aGBR) that is phase locked to the stimulus. This is referred to as stimulus entrainment. Entrainment occurs when two independent oscillatory mechanisms become coupled due the alignment in phase of their respective periods, and it is an important characteristic of interacting brain rhythms. Gamma-band synchronization is considered to increase the strength of entrainment on target neural networks, establishing exclusive neural communication links necessary for cortical computation, while the influence of gamma-band synchronization can be either general or specific depending on how it is modulated (Fries, 2009). It has been proposed that the function of gamma band activity with regard to memory processes lies in the fine temporal precision required for the induction of neural coupling and long-term potentiation to create learning-dependent changes (Axmacher et al., 2006). Thus it is hypothesised that the paradigm used primes the system with an oscillatory code, reasoning that if it is an effective code responses to target sounds will be faster depending on the harmony relationship that is expected, and

violations to these expectancies.

Music cognition concerns the cortical processes underlying our appreciation of music (Justus & Bharucha, 2002). According to Rohrmeier and Koelsch (2012) expectation and prediction (reliant upon temporally accurate protentive coding) are necessary for music perception. This research aims to apply previous findings regarding the temporal dynamics involved in cortical binding to an examination of the effects of priming on responses to violations of these expectancies. With regard to harmonic and melodic prediction, the EEG has revealed that syntactically irregular chords elicit ERP's with negative polarity and peak latencies of around 150-350 ms post stimulus onset (Rohrmeier & Koelsch, 2012). Two such predictive mechanisms are of interest to our proposed research; the early right anterior negativity (ERAN), and the right anterior temporal negativity (RATN) responses. Both are passively evoked responses to deviant stimuli elicited by music-syntax violations. However the RATN has only been observed in the cortical activity of musicians (Rohrmeier & Koelsch, 2012). Specifically our research hypothesised differences in the neural networks recruited by musicians and non-musicians during pitch processing, while revealing an increase in the strength of the RATN response in musicians due to musical training. In addition to the neural recruitment involved in this process, the behavioural evidence strongly suggests an interaction between the evoked aGBR's and a 6.69-Hz theta rhythm (possibly others) (Elliott, 2014). This research aims to establish convergence between the behavioural evidence and neural activity with regard to the involvement of neural networks in pitch processing and interactions in phase between emergent oscillatory systems and endogenous brain rhythms.

Method

Participants

Adult participants (music group ($n = 6$), ranging from beginner to Grade 8, and non-music group ($n = 6$)) were prepared for the MEG scanner.

Stimuli

A sequence of two sound stimuli (pip-trains) were used. The first was the entrainer, which carried a repeated sequence of a four pip prime; one 1,000 Hz deviant pip followed by three 500 Hz baseline pips. The second was the target stimulus which contained two different conditions—absent or present. The target absent (TA) pip-train consisted of only 1,000 Hz pips while the target present pip-train carried alternating pips which were either harmonically related to the entrainer (harmonic target present (HTP), 1000 Hz and 2000 Hz) or inharmonically related (inharmonic target present (ITP), 1000 Hz and 2400 Hz, in a ratio of 5:12). The diagram below (Fig. 1) illustrates the target conditions. Participants were instructed to respond as rapidly and accurately as possible to the presence or absence of the target tone in the second sound stimulus.

Design

The parameters of these stimuli were designed to replicate the stimuli in the paradigm which produced the previous 'pop-out' effect at 33 pips per second (pps) entrainment rate—that is slowing of reaction time (RT) responses to harmonic responses at a 33 pps

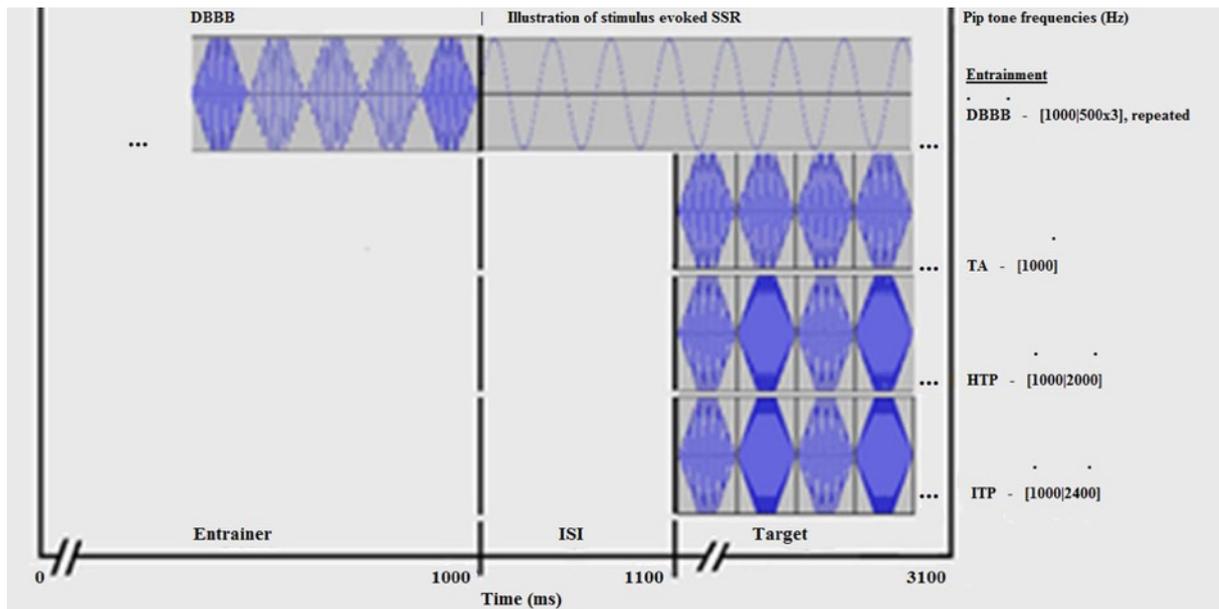


Fig. 1. Illustration of the target conditions for a 33 pips per second (pps) rate of entrainment. The pip train at the top of the diagram represents a repeated four pip entrainment sequence which is terminated at 1000 ms and comprises a total of 33 pips in this instance. The wave representation following this point is illustrative of the stimulus evoked steady state response (SSR). All three target examples are illustrated. Each is randomly selected during individual trials, and are presented following an interstimulus interval (ISI) of 100 ms.

entrainment rate, resulting in a deviation in the trend (see Aksentijevic et al. 2011, 2014, for details). This was a mixed design with a between group factor, and within group ($5 \times 2(2)$) factors rate (31, 33, 35, 37, and 39 pps), and target (TA and target present—HTP and ITP).

Magnetoencephalography Measurement

Experimental sessions (1 per participant, ~ 50 mins duration) were conducted at the Northern Ireland Brain Mapping (NIFBM) facility, at the Intelligent Systems Research Centre, Magee Campus, Ulster University, Derry, Northern Ireland. The continuous raw MEG was recorded per participant, per block (200 trials) using the 306-channel whole head MEG Elekta Neuromag system (Helsinki, Finland), comprising 204 gradiometers and 102 magnetometers. Ocular movements and cardiac activity were measured for cleaning purposes using four electrooculograph (EOG) electrodes (2 horizontal and 2 vertical), and one cardiac muscle electrode. Signals were digitised with a bandwidth of 0.1Hz to 300 Hz and a sampling rate of 1000 Hz. Sound stimuli were presented binaurally via ER-3A ABR Insert Earphones, and the decibel level was attenuated to 50 SPL, as measured by a Precision Gold (IEC 651 TYPE II) sound level meter (model #: N05CC).

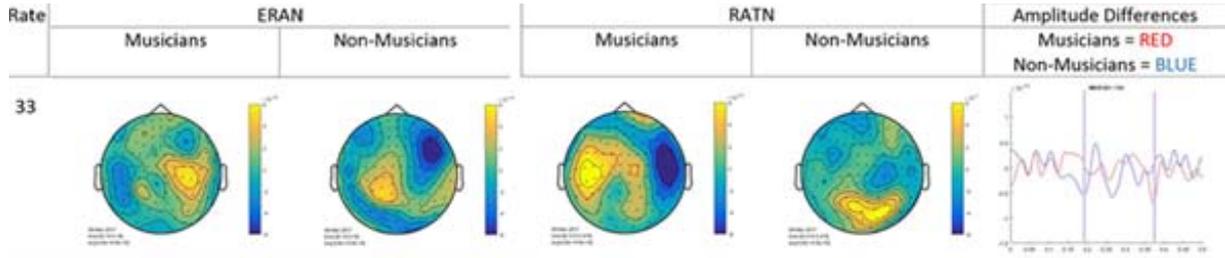


Fig. 2. Illustration of the 33 pps condition topography of the sensor level time frequency analysis for post stimulus latencies corresponding with the ERAN and RATN response windows (175-225 ms and 350-400 ms, respectively), for both groups, and the difference wave—at the MEG 1321 sensor. The difference wave (far right) is a measure of the difference in amplitude of cortical activity (event-related field’s - ERF’s) to harmonic and inharmonic stimuli from stimulus onset to 500 ms post-stimulus onset. The shaded line on the difference wave plots highlights the point in time corresponding to the topographic images preceding it.

Data Preprocessing

The continuous data were filtered off-line using Max Filter temporal signal space separation (tSSS). Data were cleaned and analysed using the FieldTrip toolbox in Matlab. Each participant’s data were resampled to a rate of 500 samples per second (500 Hz), epoched, and concatenated across blocks. Cleaning included manual, squid-jump, and muscle artifact rejection, followed by removing outliers and running an independent component analysis (ICA) to remove the remaining components resulting from eye or cardiac movements. The clean data for all participants were averaged according to condition (HTP \times Rate, and ITP \times Rate).

Results and Discussion

Data Analysis at Sensor Level

Analysis of the difference waves for each entrainment rate by group condition, was conducted using sensor MEG1321, located over the right anterior temporal region. The RATN response was only observed in the cortical activity of musicians following an entrainment rate of 33 pps. The ERAN response appeared to be most clearly present among non-musicians at this rate also (Fig. 2).

A Mann-Whitney U test did not reveal a significant group difference for either response at any rate of entrainment, although the group difference at 33 pps for the RATN response was much closer to significance [$p = 0.093$, compared to other rates ($p \geq 0.3$)]. Cluster analyses were conducted to determine significant event-related field (ERF) components during the latency windows corresponding to both the ERAN and RATN responses for each entrainment condition and group. A significant ERF was elicited within the RATN response window at 35 pps entrainment in the musician group ($p = .02$, latency of ~ 365 ms), located in the right anterior temporal region, however the ERF component was positive. In order to examine whether there were differences in the RATN response as a result of experience playing a musical instrument, musicians data were

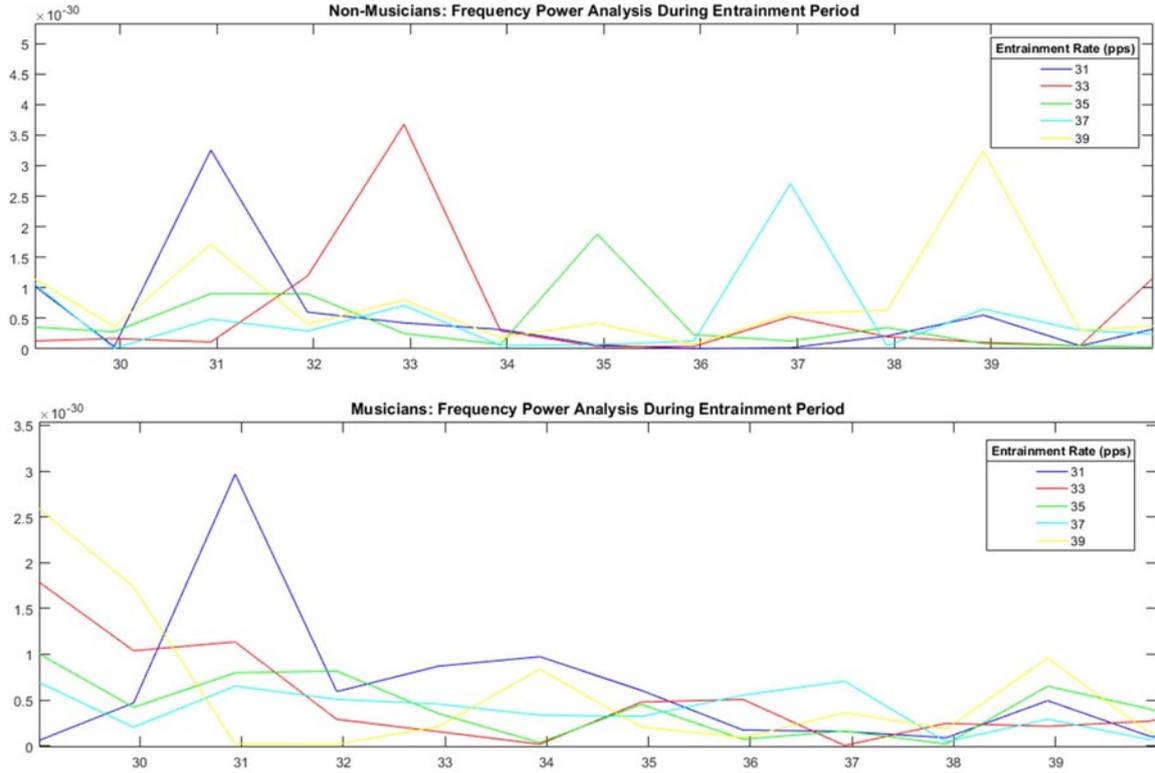


Fig. 3. Illustration of the amplitudes of frequencies between 20 and 50 Hz during the 100 ms ISI between entrainer and target, for each entrainment condition. Top: the graph for non-musicians. Bottom: the graph for musicians.

separated into two groups: grade 4 and lower ($n = 4$), and grade 8 ($n = 2$), and a cluster analysis was performed, again during the RATN response window. For both the lower grade and the grade 8 groups, at an entrainment rate of 35 pps, a positive ERF component was observed that was significant ($p = 0.0001$). For the lower grade group this is located in the anterior temporal to frontal region, with a latency window of between ~ 360 ms, while for the grade 8 group this response occurs much more precisely at a latency of 366 ms and appears to be very strong, but is located in the right frontal region. For the grade 8 group only, positive ERF components, again significant ($p = 0.001$), were observed at entrainment rates of 37 and 39 pps, at later latencies of 398 and 388 ms respectively, located in the right posterior temporal, and right frontal regions respectively.

A frequency power analysis was conducted to examine group differences in the amplitude of the evoked response to the entrainer. A low pass filter of 50 Hz and a high pass filter of 20 Hz was used on the grand averaged data for each entrainment by group condition, with a window spanning the duration of the inter-stimulus interval (ISI) (-100 to 0 ms), covering right and left temporal MEG sensors. For the non-musician group clear peaks were revealed at the evoked frequency corresponding to entrainment rate. However for the musician group the only clear peak matching the corresponding entrainment condition appears at 31 Hz for the 31 pps condition. Peaks with a much lower amplitude ($\leq 1 \times 10^{-30}$) were observed at corresponding frequencies for the 35, 37, and 39 pps entrainment conditions. However at the 33 pps rate of entrainment, there appears to be a complete absence of a 33 Hz signal (see Fig. 3).

The topography of the ERAN and RATN responses at the location defined in pre-

vious research (Maess et al., 2001), seemed to suggest that an entrainment rate of 33 pps (previously resulting in a pop-out effect) best facilitated an ERAN response in non-musicians and an RATN response in musicians (see Fig. 2). However group differences were not significant and the cluster analyses revealed that a significant response difference occurred during the 35 pps entrainment condition for the musicians only ($p = .02$), which was a right anterior temporal positive response. In support of this response, the behavioural data revealed that while musicians reaction time (RT) responses were faster in every condition compared to non-musicians, and responses to inharmonic targets were faster for both groups compared to harmonic responses—at 35 pps entrainment both groups’ responded to both targets faster than at other rates, irrespective of harmony. Therefore at the sensor level, there are differences in the cortical responses to inharmonic stimuli dependent on musical training based on the observed significant positive ERF component, within the RATN latency window—however contrary to previous research this response is positive in polarity. Interestingly, this response became much stronger, more temporally specific (366 ms latency), and more frontal in location for the musicians with grade 8, which suggests that music experience does strengthen a response to deviant stimuli with a later latency than the ERAN response. According to Mussachia et al. (2007), musicians demonstrate practice related changes, i.e. larger and earlier (10 ms) post-stimulus onset ABRs, and enhanced phase locking to stimulus periodicity (presumed to underlie pitch perception), as well as enhanced representation of the fundamental frequency (f_0). It has been suggested that this is due to a combination of top down influence and Hebbian learning. The former due to musical training modifying the neural architecture required for performance, beginning with higher processing areas and gradually enhancing lower sensory areas and the latter resulting from simultaneous activation of pre and post synaptic auditory brainstem neurons which strengthens the efficacy of the brainstem responses encoding sound. Thus music training leads to reciprocal efferent and afferent plasticity which strengthens the subcortical and cortical centres of the auditory corticofugal system.

The findings from previous research using a similar visual priming paradigm, thus far suggest a complex interplay between fast frequencies in the gamma range, and the modulating effects of slower rhythms (Elliott, 2014). Fast frequencies carry feedforward information such as stimulus input and prediction errors, and slower frequencies feedback templates and update the predictions (Michalareas et al., 2015). In this way anticipatory coding prepares the system to interact with the environment based on continuously updated object representations and predicted events. It may be that the entrainer benefits the non-musically trained brain by providing an anticipatory code in advance, while musicians are aided by well-established tone representations (templates), in higher cortical areas, and do not rely on the entrainer in the same way. The frequency power analysis provides some support for this theory, although the only entrained frequency observed was that corresponding to a 31 pps rate of entrainment, and this did not result in response facilitation. In addition the topography of the ERAN response in non-musicians and the RATN response in musicians appeared well defined at a 33 pps entrainment, while a 33 Hz frequency is almost absent during this entrainment condition. Aksentijevic et al. (2014) has argued that whereas rate-specific entrainment prepares non-musicians for target processing, in the case of a musically trained brain the entrainment frequency modulates the envelope of the carrier frequencies—the curve outlining the extremes of the oscillating signal, and at critical binding frequencies this facilitates a response overall. Given the findings of this research perhaps frequencies which are not critical for binding

can disrupt the normal binding process in the musically experienced brain.

Further analyses are required, which will include a source analysis, using forward modelling, to look at differences in the strength of ERAN and RATN responses depending on group and entrainment frequency, and a cross frequency analysis to determine if there is gamma-theta coupling (during entrainment, and following target presentation). This latter analysis will address the question regarding an interaction in phase between gamma and theta, suggested by previous research, and may also shed more light on the results from the frequency power analysis.

Acknowledgements

This research was sponsored by the Irish Research Council (IRC), and the Northern Ireland Functional Brain Mapping (NIFBM) facility, Ulster University, Magee Campus, Derry.

References

- Aksentijevic, A., Barber, P. J. & Elliott, M. A. 2011. Process timing and its relation to the coding of tonal harmony. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 1628.
- Aksentijevic, A., Smith, A. & Elliott, M. A. 2014. Rate-specific entrainment of harmonic pitch - Effects of music training. *Music Perception: An Interdisciplinary Journal*, 31, 316-322.
- Axmacher, N., Mormann, F., Fernandez, G., Elger, C. E. & Fell, J. 2006. Memory formation by neuronal synchronization. *Brain Res Rev*, 52, 170-82.
- Elliott, M. A. 2014. Atemporal equilibria: pro- and retroactive coding in the dynamics of cognitive microstructures. *Front Psychol*, 5, 990.
- Fries, P. 2009. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu Rev Neurosci*, 32, 209-24.
- Justus, T. C., & Bharucha, J. J. (2002). Music perception and cognition. In S. Yantis (Volume Ed.) and H. Pashler (Series Ed.), *Stevens' Handbook of Psychology* (Third Edition, pp. 453-492). New York: Wiley.
- Rohrmeier, M. A. & Koelsch, S. 2012. Predictive information processing in music cognition. A critical review. *Int J Psychophysiol*, 83, 164-175.
- Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J.-M., Kennedy, H. & Fries, P. 2015. Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron*, 89, 384-397.
- Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced sub-cortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, 104(40), 15894-15898.

SOME TEMPORAL CHARACTERISTICS OF MUSIC COGNITION

Mark A. Elliott, Deirdre Farrell, Caitríona Eames, and Naomi du Bois
School of Psychology, National University of Ireland, Galway, Ireland

Abstract

We investigated whether pitch relations (chords) are processed via precise temporal mechanisms, and whether the use of temporal mechanisms depends upon musical training. Musicians and non-musicians were primed with an external gamma-band entrainer, presented at rates 31, 33, 35, 37 pips-per-second (PPS). Chord sequences were concurrently presented and included target chords (either consonant or dissonant) presented in or out of phase with the ongoing entrainer. Participants were asked to respond as rapidly as possible with their judgment of the consonance or dissonance of the target chord. Musicians responded significantly faster to out-of-phase dissonant targets and non-musicians responded significantly faster to out-of-phase consonant targets. A priming bias for out of phase targets suggests that cortical mechanisms coding harmony are precisely phase-locked to stimulus onset and thus analyses of harmony occurs at a very early stage of information processing. Oscillatory gamma activity presents itself as a candidate neural mechanism for the early processing of tonal features and the prospective coding of future tonal events.

The significance of synchronized gamma activity in auditory perception was brought to attention by Galambos, Maekig and Talmaschoff (1981), who reported auditory steady state responses (ASSR) to be particularly prominent at 40-Hz (mid gamma range). This has since been demonstrated with a sequence of clicks (Tiihonen, Hari, Kaukoranta, & Kajola, 1989), tone-pips (Gutschalk, Oldermann, & Rupp, 2009), and amplitude-modulated tones (Picton, Skinner, Campagne, Kellet, & Maiste, 1987). In a recent study, Aksentijevic, Barber, and Elliott (2011) found that external oscillation entrainment in the gamma band at 33 pips per second (pps ie at 33 Hz) expedited reaction times (RTs) to harmonic targets carrying fractional multiples of the priming frequency. The induced gamma activity caused an "inharmonic popout" or increased salience of harmonic partials that were that were fractional multiples of the primed tones. Aksentijevic et al. (2011) proposed a model of harmony processing in which incoming resolved spectra are compared against internalized harmonic templates (formed via oscillatory synchronization). We assume this model to generalize to pitch processing, where pitch is determined according to the best matching templates.

Once pitch is coded, it is perceptually organized along a horizontal (e.g. melody) and vertical (eg. harmony) axis. In our musical system, pitch combinations (intervals and chords) exist in a hierarchy that determines their music-structural importance and relevance in musical piece (Lerdahl & Jackendoff, 1983). The rules of tonality specify the arrangement of chord functions within harmonic progressions (Krumhansl and Toivainen, 2001). Chords that obey the rules of tonal harmony are regarded as consonant (subjectively pleasant or stable sounding) while those that violate the rules are regarded as dissonant (subjectively harsh or requiring resolution). Most of the music composed and produced today presupposes a sensitivity to tonal music syntax and despite an inability to articulate musical intuitions, untrained listeners are nevertheless capable of making judgments about chords and their relations that are highly sophisticated and consistent with music-theoretic descriptions (Unyk & Carlson; 1987, Tillman & Lebrun-Guillard,

2006). Recent neuroimaging studies, using harmonic violations to investigate tonal processing, have suggested that the extraction of harmony involves the activity of multiple brain areas. While it is unclear whether these regions operate in isolation or interact, it is speculated that communication between these brain areas would be characterized by transient phase relationships between oscillatory activities of underlying neuronal populations (Ruiz, Koelsch & Bhattacharya, 2008).

In the current study, we hypothesized that external gamma band entrainment (see Aksentijevic et al., 2011) presented concurrently with a sequence of chords would facilitate the communication between brain areas responsible for harmony processing. We hypothesized that the neuronal mechanism involved in the processing of tonal harmony occurs at an early stage of information processing and is therefore phase-locked to the stimulus. On this basis we hypothesized that we could open/close temporal windows for communication by presenting a target in phase or out of phase alignment with the gamma prime. We expected that the presentation of targets in phase with gamma entrainment would facilitate the exchange of information between oscillating neural assemblies and that the presentation of targets out of phase with the entrainment would inhibit the exchange of information. Based on previous findings of prospective coding mechanisms (Aksentijevic et al., 2011; Elliott, 2014) and the significance of 33 Hz in Aksentijevic and Aksentijevic-like paradigms we hypothesized that maximal priming potential would occur at 33 Hz. In studies of pitch extraction, it has been observed that gamma priming is not found in musicians (Aksentijevic, Northeast, Cauty & Elliott, 2013). Given that the perceptual nuances of tonal harmony are augmented with experience (McDermott, Keebler, Micheyl, & Oxenham, 2010, Bidelman, Gandour & Krishnan, 2011), we used a broad sample of musicians and non-musicians to examine the interaction between stimulus frequency (31–37 Hz) and perceptual harmony. We expected that musicians would have a reduced reliance on external gamma entrainment.

Method

Participants

Twenty-eight participants took part in this study (12 female, mean age = 21.52 years, SD = 2.3 years); 14 were musicians, 12 of whom had received 2 to 14 years of formal training (mean = 6 years, 4 months), while 2 musicians were self-taught over a 10–14 year period. All musicians were trained on pitched instruments. Five participants were classically trained. Fourteen participants were non-musicians and reported no engagement with music beyond listening. All participants reported normal hearing and provided informed consent. The study was approved by the Ethics Committee of the Department of Psychology, National University of Ireland, Galway.

Apparatus and Stimuli

Experimental trial and stimulus generation were controlled by an IBM compatible computer. Trial presentation and data collection were implemented using Superlab (v.4, Cedrus). The stimuli were presented diotically via Beyerdynamic DT 880 Pro headphones. Participants responded to experimental stimuli on a Cedrus RB x360 response pad. Each trial consisted of a tone-pip train (prime) and overlapping chord sequence. The tone pip train was designed to generate an evoked gamma-band oscillatory steady state response

(SSR; Galambos, Makeig & Talmachoff, 1981) and was delivered in 4 frequency conditions [31, 33, 35 and 37 pps (pips per second)]. Pip duration was a reciprocal of presentation rate (e.g. 30.3 ms at 33 pps). Individual tone pips were generated in Audacity 2.0.5 and were offset and onset-ramped with symmetrical ramps that had a plateau of 33% of the overall period. The pip trains used in this study were modeled on the entrainment described in Aksentijevic et al., (2011).

A tone-pip train was presented independently at the beginning of each trial. Following 17–19 tone pips (approx. 515 ms), a sequence of chords was introduced and presented over the remaining pip train. The timing of sequence onset (17–19 pips) varied between rate conditions to ensure that the initial chord was optimally aligned with the phase of the entrainer. All entrainers and chord sequences carried 500 Hz and were presented at the intensities of 61 dB and 64 dB, respectively. The adjustment of intensity was carried out to control for the potentially distracting nature of the entrainer on task performance. Two chord sequence-types were presented: chord progressions that concluded on a consonant chord and progressions that concluded on a dissonant. The task was presented as a discrimination exercise for which participants were asked to respond to consonant and dissonant chord closures. Both stimulus-types consisted of the same chord progression and differed only with respect to their terminating chords.

Sequences varied in length from four to five chords. Consequently, there were four categories of chord sequences: a short sequence with a consonant closure, a longer sequence with a consonant closure, a short sequence with a dissonant closure, a longer sequence with a dissonant closure. This experimental manipulation was carried out to offset the repetitive nature of the task thus controlling for possible incidental entrainment of brain activity related to attentional deployment, in the EEG alpha band (i.e. 8–12 Hz). Both targets (consonant and dissonant closures) were presented an equal number of times. Consonant terminal chords were Tonic chords (I). The harmonic progression in the consonant condition was dominant to tonic (V–I). This is known as a perfect cadence and is the most harmonically appropriate progression at the end of a sequence. A perfect cadence commonly serves as the final event in a musical piece in that it elicits a feeling of closure and completion.

Dissonant terminal chords were Neapolitan sixth (N^6) chords. Due to its unusual half step root relationship to the tonic and overall distant relationship to the established musical key, the Neapolitan is perceived as striking and dissonant. The harmonic progression in the dissonant progression was dominant to Neapolitan (V– N^6). This is a highly inappropriate progression as the presentation of the dominant sets up the expectancy for a tonic. This progression therefore elicits a high degree of dissonance and surprise. Individual chords were created and exported as WAV files using composition and notation software MuseScore 2. All chords were set to acoustic piano timbre to mimic an authentic musical stimulus. The chords (WAV form) were imported into Audacity 2.0.5 and inaudible fade-in/fade-out signals were trimmed from the audio. Individual chords were superimposed onto pip trains (prime) in accordance with the phase alignment of tone pips. Each chord initiated at the onset of a tone pip and terminated at the offset of another tone pip).

Chords lasted roughly 250 ms, however, they were varied in length (i.e. tone pip count) to control for the entrainment of an extraneous rhythm induced by the temporal regularity of chord onset. In all trials, the terminal chord (target) was longer and was ramped at offset to allow for the sound to resonate (aiding recognition of target) and to prevent a constrictive release. Tone pips discontinued midway through the terminal chord,

again to prevent the plosive effect of sudden pip release. Target (terminal) chords were placed precisely out of phase alignment with the phase of tone-pip presentations on 50% of trials and were placed in-phase on the other 50% of trials. This was intended to offset the gamma-band oscillatory steady state response (SSR) achieved by the entrainment.

Design and Procedure

The study employed a within-subjects design with targets (consonant closure, dissonant closure) and factors rate (31, 33, 35, 37 pps) and alignment (in phase/out of phase), with the between-subjects factors musician and non-musician. There were 40 trials per condition resulting in 640 trials per participant. Trials were presented in eight blocks of 80 trials per block. The experiment was conducted in a sound-attenuated experimental cabin in a laboratory suite housed in the School of Psychology at NUI Galway. Stimulus intensity was held constant throughout [average 60 dB sound pressure level (SPL)], determined using an Aadastra analog sound level meter. All stimulus sequences were presented pseudo-randomly on a session wise basis. Each trial commenced with a blank screen. A randomized interval between two adjacent trials was set between 400 and 500 ms and was followed by stimulus presentation. Response time was measured from target offset and a new trial was initiated automatically from the response.

Magnetoencephalography Measurement

Experimental sessions (1 per participant, \sim 50 mins duration) were conducted at the Northern Ireland Brain Mapping (NIFBM) facility, at the Intelligent Systems Research Centre, Magee Campus, Ulster University, Derry, Northern Ireland. The continuous raw MEG was recorded per participant, per block (200 trials) using the 306-channel whole head MEG Elekta Neuromag system (Helsinki, Finland), comprising 204 gradiometers and 102 magnetometers. Ocular movements and cardiac activity were measured for cleaning purposes using four electrooculograph (EOG) electrodes (2 horizontal and 2 vertical), and one cardiac muscle electrode. Signals were digitised with a bandwidth of 0.1Hz to 300 Hz and a sampling rate of 1000 Hz. Sound stimuli were presented binaurally via ER-3A ABR Insert Earphones, and the decibel level was attenuated to 50 SPL, as measured by a Precision Gold (IEC 651 TYPE II) sound level meter (model #: N05CC).

Results

Participants performed the task with high accuracy (96% correct for consonant and dissonant targets). Trials with error responses were removed from the data prior to analysis. Error RTs were slower than correct RTs and an analysis of the probability correct by RT showed no significant correlation between RT and accuracy. Examination of the correct RTs revealed non-normal distribution with a pronounced positive skew. RT distributions were approximately lognormal on a Kolmogorov ‘D’ test and as a result, subsequent analyses were performed on the exponents of the means of log-transformed RT distributions.

The RT means were subject to a three-way repeated measures analysis of variance (rANOVA) with the within-subjects factors rate (31, 33, 35, 37 pps), synchrony (in-/out-of phase) and harmony (consonant, dissonant) and the between-subjects factor musicianship (musicians, non-musicians). Violations of the homogeneity of variance assumption were corrected by applying Greenhouse Geisser or Huynh-Feldt adjustment criteria. There

was a significant main effect for harmony [$F(1, 26) = 10.50, MSE = 5906, p < .005, \eta^2_p = .29$] with consonant targets being registered more rapidly than dissonant targets [mean difference = 23.5 ms, standard error of the mean (SEM) = 7.26 ms]. The main effect of synchrony was significant [$F(2, 26) = 7.11, MSE = 5906, p < .05, \eta^2_p = .26$] with faster RTs to targets that were presented out of phase alignment with the entrainer relative to those presented in phase (mean difference = 7.18 ms, SEM = 2.69 ms). There was a main effect for rate [$F(3, 78) = 6.48, MSE = 832, p = .001, \eta^2_p = .20$], reflecting a tendency for RTs to be significantly slower when primed at 33 PPS. The effect of rate was influenced by synchrony as indicated by a significant rate by synchrony interaction [$F(3, 74.63) = 8.71, MSE = 731, p < .001, \eta^2_p = .25$], resulting from a significant difference between RTs to targets presented in/out of phase alignment at 35 PPS (Mean difference = 13 ms, SEM = 2.65, $p < .001$) and 37 PPS (Mean difference = 23 ms, SEM = 4.64, $p < .001$) and other rates, 31 PPS (Mean difference = 9 ms, SEM = 6.55, $p > .5$) and 33 PPS (Mean difference = 9 ms, SEM = 5.70, $p > .5$). There was also a significant interaction between rate and harmony [$F(3, 76.83) = 3.7, MSE = 703, p < .001, \eta^2_p = .11$], resulting from a significant difference in RT when presented with consonant/dissonant targets at 31 PPS (Mean difference = 34 ms, SEM = 4.64, $p < .005$), 35 PPS (Mean difference = 27 ms, SEM = 6.57, $p < .001$) and 37 PPS (Mean difference = 22 ms, SEM = 6.80, $p < .005$).

There was a significant three-way interaction between synchrony, harmony and musicianship [$F(1, 26) = 4.23, MSE = 437, p = .05, \eta^2_p = .14$]. Post-hoc analysis was carried out using a rANOVA with factors harmony, synchrony (collapsed over frequency) and musicianship. This analyses revealed the significant interaction to result from the combination of a significant RT advantage for musicians when presented with dissonant targets out-of-phase with the onset of entrainers (Mean difference = 8.76 ms, SEM = 3.40, $p < .05$) and an RT advantage for non-musicians when presented with consonant targets, out-of-phase with the onset of entrainers (Mean difference = 13.73 ms, SEM = 5.75, $p < .05$). Overall, there was no main effect, nor were there any other interactions of any other factors with musicianship. This indicates that musicians did not process any other stimulus characteristics differently to the non-musicians.

Discussion

Data Analysis at Sensor Level

A principle finding was of a general RT facilitation to targets that were presented out of phase with the entrainer. This effect could not be mediated strategically as participants were unable to report the phase synchrony of the prime. It suggests that neuronal temporal coding at gamma frequencies that are phase-locked to stimulus activity, mediates the extraction of perceptual harmony. We suggest that synchronized gamma oscillations activate the cortical neurons coding a given pitch and that these oscillations are extended to neuronal populations of related pitches. By extension, synchronized gamma activity may present a candidate neural mechanism for the generation of the total tonal experience (through phasic stimulation from pitch-chord-key units), as described in connectionist models of tonality (see Bharucha, 1987).

Contrary to hypothesis, a RT advantage to out-of-phase targets indicates that maximal priming was not achieved when targets were optimally aligned. Thus this condition did not facilitate the exchange of information between oscillating neural assemblies as expected. A general RT facilitation was observed for consonant targets. This could

be attributable to the gamma entrainment or to the tonal context established by the harmonic progression (see Bharucha & Stoeckig, 1987). However, a baseline investigation of gamma entrainment vs. no entrainment may be necessary to verify this.

While there was no difference in overall RTs between non-musicians and musicians, musicians detected out-of-phase targets more efficiently when they were dissonant, non-musicians detected out-of-phase targets more efficiently when they were consonant. This indicates that the temporal dynamics of harmony processing differs between the two groups and supports recent evidence of differential gamma power (indicative of the number of synchronized neurons) in response to musical stimuli in musicians as compared with non-musicians (Bhattacharya, & Petsche 2001; Roberts, 2007). Given evidence of more robust and synchronous gamma activity in musicians, we expected that musicians would have an decreased sensitivity to entrainment at gamma-band frequencies and that consequentially, no effect would be observed. Contrary to this hypothesis, musicians RTs were sensitive to half phase manipulations in the prime, indicating that the entrainment did modulate the synchrony of ongoing cortical oscillations in musicians. The RT advantage for out-of-phase dissonant targets in musicians could be due to the function both of gamma oscillations as an anticipatory mechanism (Elliott, 2014), and the schematic discrepancy of the dissonant chord. If enhanced gamma activity reflects both the expectation and realization of predicted events, musicians may be more sensitive to the perceptual nuances of musical syntax and therefore experience a higher degree of schematic discrepancy when a syntactically unexpected event occurs. The harmonic progression of a dominant chord function to a Neapolitan chord (V–N⁶) is extremely rare in modern composition and is therefore highly inconsistent with a (western) musician’s schema of tonal relations.

References

- Aksentijevic, A., Barber, P. J., & Elliott, M. A. (2011). Process timing and its relation to the coding of tonal harmony. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1628–1642.
- Aksentijevic, A., Northeast, A., Canty, D., & Elliott, M. A. (2013). The oscillatory entrainment of virtual pitch perception. *Frontiers in Psychology*, 4.
- Bharucha, J. J (1987). MUSACT: A Connectionist Model of Musical Harmony. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*. Hillsdale,
- Bharucha, J. J., & Stoeckig, K. (1987). Priming of chords: Spreading activation or overlapping frequency spectra?. *Perception & Psychophysics*, 41(6), 519–524.
- Bhattacharya, J., & Petsche, H. (2001). Musicians and the gamma band: a secret affair? *NeuroReport*, 12(2), 371–374.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition*, 77(1), 1–10.
- Elliott, M.A. (2014). Atemporal equilibria: pro- and retroactive coding in the dynamics of cognitive microstructures. *Frontiers in Psychology*, 5:990. doi:10.3389/fpsyg.2014.00990
- Galambos, R., Makeig, S., & Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proceedings of the National Academy of Sciences, USA*, 78, 2643–2647.
- Gutschalk A, Oldermann K, Rupp A (2009) Rate perception and the auditory 40-Hz steady-state fields evoked by two-tone sequences. *Hear. Res.*, 257:83–92.
- Krumhansl, C. L., & Toiviainen, P. (2001). Tonal cognition. *Annals of the New York*

- Academy of Sciences*, 930(1), 77–91.
- Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception: An Interdisciplinary Journal*, 1(2), 229–252.
- McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *The Journal of the Acoustical Society of America*, 128(4), 1943–1951.
- Picton, T. W., Skinner, C. R., Campagne, S. C., Kellet, A. J. C., & Maiste, A. C. (1987). Potentials evoked by the sinusoidal modulation of the amplitude and frequency of a tone. *Journal of the Acoustical Society of America* 82, 175–178.
- Ruiz, M. H., Koelsch, S., & Bhattacharya, J. (2009). Decrease in early right alpha band phase synchronization and late gamma band oscillations in processing syntax in music. *Human brain mapping*, 30(4), 1207–1225.
- Tiihonen, J., Hari, R., Kaukoranta, E., & Kajola, M. (1989). Interaural interaction in the human auditory cortex. *Audiology*, 28, 37–48.
- Tillman, B. & Lebrun-Guillard, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgements of chord sequences. *Psychological Research*, 70, 345–358.
- Unyk, A. J., & J. C., Carlson (1987). The influence of expectancy on melodic perception. *Psychomusicology*, 7, 3–23