The effect of musical training on auditory grouping

Sarah Sauvé
Department of Psychology, Goldsmiths University of London
sas703@gold.ac.uk

Lauren Stewart
Department of Psychology, Goldsmiths University of London
l.stewart@gold.ac.uk

Marcus Pearce
Department of Engineering and Computer Science, Queen Mary University of London
marcus.pearce@qmul.ac.uk

In: Jakubowski, K., Farrugia, N., Floridou, G.A., & Gagen, J. (Eds.)
Proceedings of the 7th International Conference of Students of Systematic Musicology (SysMus14)
London, UK, 18-20 September 2014, http://www.musicmindbrain.com/#!sysmus-2014/cfmp

Background. Auditory streaming is a process highly relevant to analyzing everyday sound environments, particularly with respect to timbre. The phenomenon of auditory streaming has a history of being studied in terms of Gestalt principles (Bregman, 1990), of pitch (van Noorden, 1975), of tempo (Bregman & Campbell, 1971; van Noorden, 1975), of timbre (Bregman & Pinker, 1978; Marozeau et al., 2013), and of attention (Botte et al., 1997; Carlyon & Cusack, 2001). All of these parameters influence the extent of auditory streaming in various ways. An increase in performance in many types of auditory tasks is seen in musicians, including streaming (Zendel & Alain, 2008), presumably a result of training and brain plasticity.

Aims. This experiment seeks to corroborate this observed effect of musical training, and further define the effects of training on specific instrument. Another goal of this experiment is to clearly demonstrate the influence of attention on streaming.

Method. In testing both non-musicians and musicians trained on specific instruments in a simple ABA-paradigm where timbre is manipulated (similar timbres presumably making streaming more difficult (Singh & Bregman, 1978; Hartmann & Johnson, 1991; Iverson et al., 1995)), we can find and analyze the *fission* and *temporal coherence boundaries* between groups. Participants will be exposed to trials via Max/MSP, and responses will be collected in the same patch. A manipulation of instructions to participants will evaluate the influence of attention on streaming: they will be instructed to hold on to either the galloping rhythm (integration) or the 2:1 rhythm (streaming).

Results. This experiment corroborates the previously observed effect of musical training on perception, demonstrated by different threshold profiles between musicians and non-musicians. It also clearly demonstrates an influence of attention on streaming while suggesting further effects of training on specific instruments. The manipulation of attention formed two boundaries, identified as the fission boundary and the temporal coherence boundary. These boundaries were significantly different between musicians and non-musicians and additionally affected by specific timbres.

Auditory streaming is the perceptual breaking apart of sound input into its component sources. It has been investigated in the context of numerous sound attributes such as pitch (van Noorden, 1975), location (Jones & Macken, 1995), periodicity (Vliegen, Moore, & Oxenham, 1999) and timbre (Iverson, 1995) among others. The role of musical training has also been extensively studied in the context of auditory skills, including auditory streaming (Zendel & Alain, 2009). As a result of training, musicians are more sensitive to changes in auditory stimuli based on pitch, time and loudness for example (Marozeau, Innes-Brown, & Blamey, 2013; Marozeau, Innes-Brown, Grayden, Burkitt, & Blamey, 2010), with discrimination thresholds being lower in musicians than in non-musicians. One problem with treating musicians as one single category is that fine differences between instrumentalists may be missed. Pantev and colleagues (Pantev, Roberts, Schulz, Engelien, & Ross, 2001) found that certain instrumentalists were more sensitive to the timbre of their own instrument than to others, as measured by auditory evoked fields. However, this has not yet been observed in auditory streaming, where an effect would be seen by a change in streaming threshold; presumably, it would take less time to detect two separate auditory objects when one's own instrumental timbre is one of these objects.

Over the last few decades, the emerging concept of auditory streaming is that of a two-staged process: the first being pre-attentive, peripheral and following Gestalt principles (Bregman, 1990; Hartmann & Johnson, 1991) and the second post-attentive, cortical and involving top-down, heuristic and schematic mechanisms (Alain, Arnott, & Picton, 2001). Though there is still ongoing debate about the role of attention in streaming, specifically whether steaming is a pre- or post-attentive phenomenon, it is clear that attention influences perception (van Noorden, 1975). With accumulating evidence for a two-staged process, a better question now is: what is the influence of attention at each stage?

Following the rules of Gestalt psychology, auditory streaming displays multi-stability. Within certain criteria, listeners can hear either coherence or segregation, based on top-down processes such as training, attention or expectation (Bendixen, Denham, Gyimesi, & Winkler, 2010). van Noorden (1975) defined two boundaries within which bi-stability was possible by manipulating attentional focus. These boundaries are the fission boundary, the difference in pitch below which segregation into two auditory objects, or streams, is not possible, and the temporal coherence boundary, the difference in pitch above which coherence into one auditory object, or stream, is no longer possible.

This research has three objectives. The first is to reveal the fission and temporal coherence boundaries in streaming by timbre through a manipulation of attentional focus. The second is to test whether these boundaries are influenced by musical training, with boundaries hypothesized to be lower for musicians. The third is to find behavioural evidence for an increased sensitivity in streaming the instrument(s) which a musician plays, expected to be shown by lower boundaries for piano sounds in pianists for example.

Method

Participants

Participants were 29 university students, 19 females and 10 males, average age 26.75 (SD = 7.88, range 18-50). Using Gold-MSI scores (D. Müllensiefen, Gingras, Musil, & Stewart, 2014), 11 participants were classified as non-musicians (score<17) and 9 as musicians (score>39). The remaining 9 were classified as amateur musicians (26<score<38). The Gold-MSI was designed as a musical sophistication test that would be non-Western biased and would recognize musical skills outside of performance such listening and writing about music, skills possessed by people such as music producers and critics. These boundaries were taken from the data, where the Gold-MSI scores were fairly evenly distributed in two chunks: below 17 and above 26. 11 participants (9 included in data analysis) scored below 17; therefore, to form equal groups for comparison, the top 9 scores were classified as musicians and equated to a score above 39. Of the amateur and top score musicians, four named the piano as their primary instrument, four named the violin, one the trumpet and the remaining nine named a range of other instruments, including voice.

Stimuli

All six timbral sounds (piano, violin, trumpet, trombone, clarinet, bassoon) were chosen from the MUMS library (Opolko & Wapnick, 2006) to be of equal pitch (A220) and were adjusted to equal perceptual length of 100ms and equal loudness, based on the softest sound. A 10ms fade out was applied to each timbral sound. All edits were done in Audacity and the final product was exported as a wave file.

The 'standard' sequence was presented at a rate of onset of 220ms and did not change within a trial. The 'target' sequence was presented at a rate of onset of 440ms, beginning 110ms after the 'standard' sequence to create the well-known galloping ABA pattern (van Noorden, 1975). The target sequence was a 30s cross-fade between the standard timbre and the target timbre. The target sequence cross-faded from standard to target timbre in the ascending condition, creating a galloping to even rhythm change, and from target to standard timbre in the descending condition, creating an even to galloping rhythm change (Figure 1). Each trial ended when the participant indicated a change in perception.

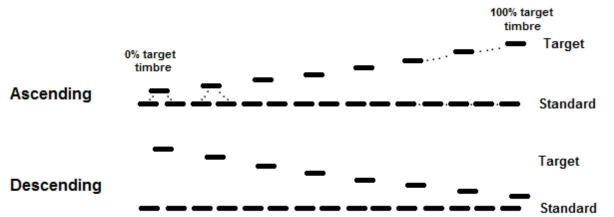
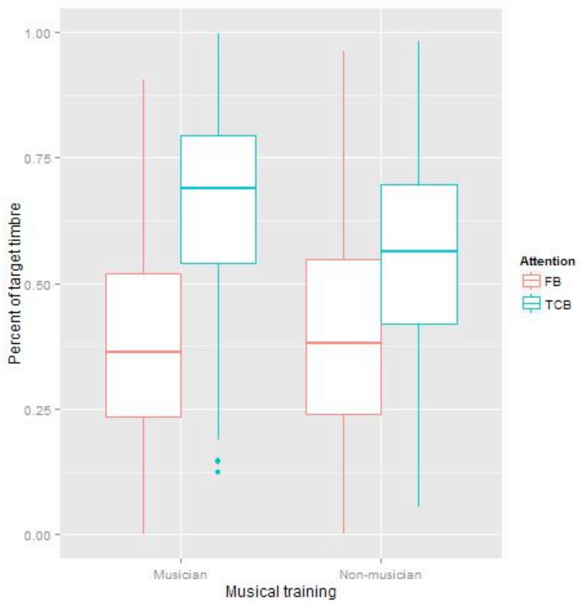


Figure 1. Illustrated stimulus for an ascending trial and a descending trial, with time represented horizontally



and timbre represented vertically.

Figure 2. : Percent of target timbre as a function of musical training and attention, with a clear division between FB and TCB. The error bars show the first and third quartiles.

Procedure

The experiment was coded and run in Max/MSP, with output heard through headphones and input taken from mouse clicks. Participants, sitting in a sound-proof booth, were first presented with a practice patch with instructions and an opportunity to listen to each timbre and rhythm separately. Up to twelve practice trials were included in the patch and questions were welcomed.

For each trial, participants indicated by clicking a button on the screen at which point the galloping sequence becomes perceived as two separate streams of standard and target tones, or the opposite for decreasing presentation. This point was recorded as percent of target timbre present in the cross-fade at that time. Each trial lasted a maximum of 30s, at which point the trial ended itself and recorded a value of '-1' for that trial, indicating that the participant had never reached a change in perception. The manipulation of attention was as follows: in two blocks, the instructions were to indicate a change in rhythm as soon as it was perceived and for the other two blocks the instructions were to hold on to the original rhythm as long as possible. Each pair of blocks contained an ascending and descending block, for a total of four blocks. For every block, every timbre modulated to every other timbre once for a total of 30 trials (6 timbres each modulating to the 5 other timbres), each separated by 4s. Participants were assigned to different orders (four possibilities) in rotation to prevent order effects.

Once all blocks were completed, participants filled out the musical training sub-scale of the Goldsmiths Musical Sophistication Index (Daniel Müllensiefen, Gingras, Musil, & Stewart, 2012; D. Müllensiefen et al., 2014).

Results

Two non-musicians' data were removed from analysis as the task was clearly not understood, leaving 9 non-musicians. Also, trials where participants did not reach a change in perception were removed from the main data and analyzed separately. The dependent variable was the percentage of target timbre present in the cross-fade of the target sequence at the time the participant indicated a change in perception, expressed in decimal value. A between-subjects ANOVA for order effects (4 levels) was non-significant ($p > .05$), as expected. An independent-samples t-test for direction (ascending/descending) was non-significant ($p > .05$), also as expected. A mixed ANOVA with musical training (2 levels: musician, non-musician) as a between-subjects variable and attentional focus (2 levels) as a within-subjects variable yielded a main effect of attentional focus, F (1, 16) = 44.94, $p < .01$ and a significant interaction, F (1, 16) = 4.61, $p = .04$ (Figure 2) with attentional focus having a stronger effect on musicians. There was no significant main effect for musical training. As attentional focus was significant, demonstrating the fission

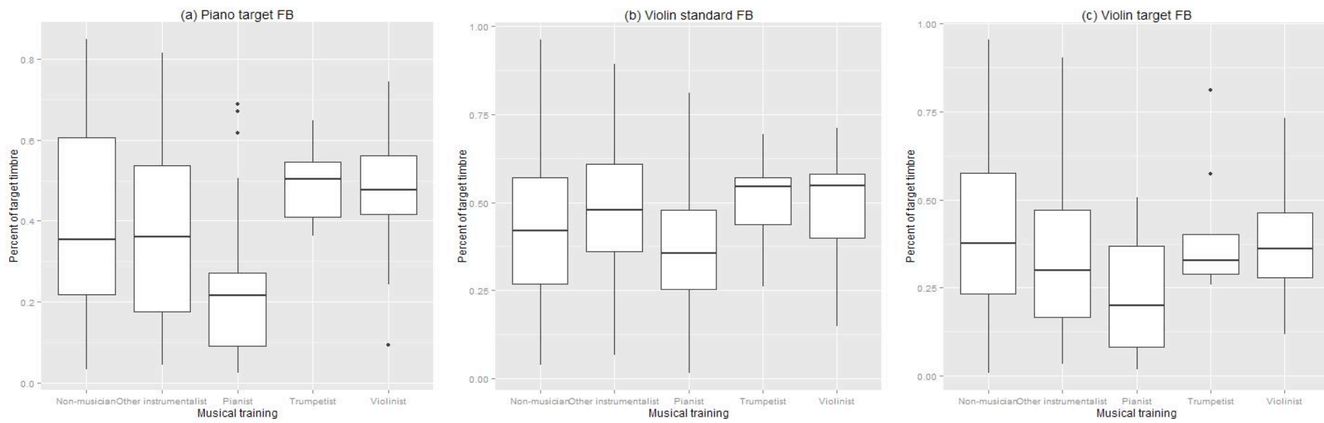


Figure 3. *: Percent of target timbre as a function of musical training when piano was the target timbre (a), when violin was the standard timbre (b) and when violin was the target timbre (c). In all cases, pianists perform best though to varying degrees. The error bars show the first and third quartiles.*

boundary (FB) and temporal coherence boundary (TCB) as separate phenomena, further analysis on the effects of musical training on the timbre discrimination threshold will be conducted for the FB only, as timbral discrimination sensitivity will be most salient at this point. For all trials where piano was the standard timbre, and all trials where piano was the target timbre, a between-subjects ANOVA for musical training

Similarly, for all trials where violin was both the standard and the target timbre, a between-subjects ANOVA for musical training (5 levels) was conducted. In all cases, Levene's test of homogeneity of variance failed; therefore pairwise comparisons with Bonferroni correction were applied. The FB was significantly lower in pianists as compared to violinists when violin was the standard, $p = .03$ (Figure 3b) and lower for pianists as compared to violinists and non-musicians when violin was the target timbre, $p = .04$, and $p < .01$, respectively (Figure 3c).

Performance was also analyzed by type of instrumentalist. An ANOVA for standard timbre (6 levels) was conducted on all pianists, with no significant effect of standard timbre. A pairwise t-test with Bonferroni correction only revealed significant difference between violin and trombone, $p = 0.04$. An ANOVA for target timbre (6 levels) was also conducted, revealing a significant effect of target timbre, after sphericity correction, F (5, 15) = 6.24, $p < .01$.

The same analysis was performed for violinists, where an ANOVA for standard

(5 levels: pianist, violinist, trumpeter, other musician and non-musician) was conducted with non-significant results for both ($p > .05$); however, both tests failed Levene's test for homogeneity of variance. Therefore, a post-hoc pairwise t-test with Bonferroni correction was conducted revealing a significant difference between pianists and every other category of musical training, $p < .05$ (Figure 3a) when piano was the target timbre.

timbre was non-significant, while Bonferroni corrected pairwise t-tests revealed several significantly differing pairs of timbres, summarized in Table 1. An ANOVA for target timbre revealed a main effect of timbre, F (5, 15) = 13.53, $p < .01$, with significant pairwise comparisons also summarized in Table 1.

Thresholds for all possible pairs of timbres were compared to see if performance was better for any particular timbral pairing; results were non-significant.

The trials where participants never reached a change in perception were separately analyzed by counting the number of instances by variable. First, the count was much higher for the TCB than the FB, as expected, $\chi 2$ (1) = 19.61, $p < .01$. The number of instances was not significantly different across different standard or target timbres but did happen for every possible timbral pair. The occurrence of timbral pairs was not evenly spread, $\chi 2$ (14) = 25.43, $p = .03$, with perception not changing most often for piano-violin, piano-clarinet, piano-bassoon and violin-trombone pairs of timbre.

A closer look at these data by type of musical training revealed some interesting trends. First, there was most often a lack of change in perception in pianists. Proportionally, there were 19.25 cases per pianist, 2 cases per trumpeter, 6.5 per violinist, 6.3 per other musician and 10.5 per non-musician. Looking at pianists more closely, this lack of change occurs equally as often for the TCB in the ascending direction and the FB in the descending direction ($\chi 2$ (1) = .63, p = .42), both understandable as these are the conditions that might cause overshooting. It also happened more when piano was the standard timbre, though not significantly ($\chi 2$ (5) = 9.1, p = .1). No effect of training appeared in pianists for the target timbre. A similar analysis for violinists revealed no significant differences between standard or target timbres, or thresholds.

p-value	Timbre pairs – standard	Timbre pairs - target
< **.01**	*Piano*-clarinet *Trumpet*-clarinet *Trombone*-clarinet *Bassoon*-clarinet	
< **.05**	*Piano*-violin *Trombone*-violin	Piano-*violin* Trombone-*violin* Piano-*clarinet* Trombone-*clarinet* *Violin*-bassoon *Clarinet*-bassoon *(p = .05)*

Table 1. Summary of significantly different timbre discrimination threshold pairs when violinists' performance was analyzed as a function of standard and target timbre. Italicized instruments are those with the lower threshold; notice a mirroring trend between the standard and target regarding the violin and clarinet.

Mirroring the main analysis, cases were also analyzed by timbre. For all cases of a lack of change in perception where piano was the standard timbre, pianists had more cases than violinists and trumpeters, $\chi 2$ (4) = 29.04, p < .01. Interestingly, cases were roughly equal for pianists and non-musicians. The same pattern can be observed when piano was the target timbre, $\chi 2$ (4) = 29.25, p < .01 though there are fewer cases overall: 14 cases of pianists not hearing a change in perception when piano was the target and 22 when piano was the standard. When violin was the standard or target timbre, a similar profile was observed: pianists tended to not reach a change in perception more often than violinists and trumpeters (p < .01 for both standard and target timbre) and non-musicians performed comparatively to pianists.

Discussion

The results have demonstrated that shifting the focus of attention has a clear effect on perception. Instructions to indicate the first hint of a change in rhythm or to hold on to the original rhythm as long as possible translate into focusing on the galloping rhythm (integration), which defines the temporal coherence boundary, or the even rhythm (segregation), which defines the fission boundary. This contradicts Carlyon and colleagues (Carlyon, Cusack, Foxton, & Robertson, 2001), who questioned the existence of the fission boundary while finding a clear temporal coherence boundary in a paper where streaming was measured outside the focus of attention. Perhaps if attention is focused away from the auditory scene, there are no separate boundaries but only streaming or lack thereof. This interpretation suggests that streaming can and does occur outside the focus of attention but is treated with finer detail inside the focus of attention, reflecting differing functions between the proposed pre- and post-attentive stages of auditory scene analysis (Snyder & Alain, 2007).

The results also show a difference in performance between musicians and non-musicians, though the observed effect is not as predicted. While discrimination boundaries were expected to be lower in musicians, the fission boundary is equal and the temporal coherence boundary significantly higher. This equates to a larger range of bi-stability, perhaps due to top-down processes having a stronger influence in musicians. Expectation can play a large role in influencing perception (Summerfield & Egner, 2009), especially with the use of directive instructions (van Noorden, 1975), and this may be stronger for musical concepts in those with musical training. One theory of auditory streaming states that coherence is the default percept, with stream segregation only occurring with accumulated supporting evidence (Bregman, 1990); however, if this was the case the descending blocks of this design would not have resulted in any change in rhythm in any trial, demonstrating the strength of expectation. Instead, it seems likely that the expectation of hearing two streams and an even rhythm allows immediate perception of segregation. Similarly, musicians have a firmer schema for rhythm and timbre and could hold coherence over larger differences when instructed to. To implicate expectation and its influence with certainty, it would be interesting to compare responses to a simple one/two stream judgment task between the current set of instructions and the absence of leading instructions, similarly to Deike and colleagues (Deike, Heil, Böckmann-Barthel, & Brechmann, 2012).

The data do not clearly demonstrate an advantage of musical training on a specific instrument to detecting specific timbre; however, some trends can be observed. The first deals with pianists: pianists tend to have lower thresholds overall, regardless of timbre analyzed. Some finer differences can also be observed in pianists: piano and violin timbres were least well discriminated when they were the standard timbre and better when they were the target timbre, though not significantly. Why pianists' timbral sensitivity includes a second instrument is an intriguing question, especially since a similar phenomenon is observed when violinists' performance is analyzed by timbre. For violinists, violin and clarinet timbres are discriminated least well when they are the standard, and best when they are the target. While there is no obvious reason why instrumentalists demonstrate dual timbre similarity here, there can be an explanation for the mirror effect between performance on standard and target timbres. If pianists are more sensitive to piano timbre, then when a trial contains the piano timbre throughout, it stands out more and will mask the incoming target timbre for longer than any other standard would, causing an increase in discrimination threshold. On the other hand, when piano enters as a new timbre, a pianist will detect it more quickly. The same can be applied to violinists. In fact, following the above logic, effects are generally stronger for target timbres.

This argument is further supported by the analysis of cases of a lack of change in perception. Though most cases of over-shooting were in the direction of the TCB, indicating that maximally different timbres can still be integrated, curiously, there were also a few cases where a change in rhythm was not detected even when timbres were identical and should have caused complete integration. It is an open question as to why this might be. When looking at these cases, there is once again an effect of piano and pianists. First, piano was involved in most of the common timbral pairs associated with a lack of change in perception. Second, pianists displayed a lack of change in perception most often in general and also more than other instrumentalists when piano was the standard and target timbre. Curiously, pianists also demonstrated this trend when violin was the standard and target timbre. There were more cases of lack of change in perception when piano was the standard timbre for pianists, supporting the above argument that when a certain timbre is the standard, it masks a new entering timbre and therefore integration can last longer for that instrumentalist.

Though no clear picture of the effect of musical training on grouping by timbre was extracted from this set of data, it is the hope that details will come out more strongly when comparing performance on the pairs of brass and woodwind instruments against brass and woodwind players, as these pairs of instruments are both closer in timbral space (McAdams, Winsberg, Donnadieu, & De Soete, 1995) than piano and violin. For this experiment, no particular timbral pair statistically stood out as easier to discriminate than another; however, in informal discussion following the experiment, several participants reported certain timbral pairings to be much easier to segregate than others (for example piano-violin as opposed to trumpet-trombone).

As a final suggestion to explain the observed data seen and discussed above and incorporating the two-stage understanding of streaming processes, perhaps the FB reflects bottom-up processing and the TCB, top-down processing. The FB is more dependent upon absolute hearing abilities: a change in timbre is detected or not and therefore could be taken as a behavioural index of bottom-up timbral segregation, which may change with instrumental training received. The TCB is susceptible to modulation by top-down schemas and expectations, where tolerance for bi-stability may increase or decrease with musical training. Further work is needed to support or contradict this proposition.

In conclusion, this research has demonstrated that when attention is focused on the auditory stimulus, the streaming process can be influenced by manipulating the focus of attention. It has shown that musicians and non-musicians respond to this manipulation differently, with musicians having a wider range of acceptance for bi-stability. Recruiting brass and woodwind players is the next step in investigating the influence of timbre-specific training. Finally, it has also made clear that streaming cannot be discussed separately from attention and expectation, whose roles and influence remain to be specified.

References

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology-Human Perception and Performance, 27*(5), 1072-1089. doi: 10.1037//0096-1523.27.5.1072

Bendixen, A., Denham, S. L., Gyimesi, K., & Winkler, I. (2010). Regular patterns stabilize auditory streams. *Journal of the Acoustical Society of America, 128*(6), 3658-3666. doi: 10.1121/1.3500695

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge: MIT Press.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology-Human Perception and Performance, 27*(1), 115-127. doi: 10.1037//0096-1523.27.1.115

Deike, S., Heil, P., Böckmann-Barthel, M., & Brechmann, A. (2012). The Build-up of Auditory Stream Segregation: A Different Perspective. *Frontiers in Psychology, 3*(Article 461), 7.

Hartmann, W. M., & Johnson, D. (1991). STREAM SEGREGATION AND PERIPHERAL CHANNELING. *Music Perception, 9*(2), 155-184.

Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance, 21*(4), 751-763. doi: 10.1037/0096-1523.21.4.751

Jones, D. M., & Macken, W. J. (1995). Organizational factors in the effect of irrelevant speech: The role of spatial location and timing. *Memory & Cognition, 23*(2), 192-200. doi: 10.3758/BF03197221

Marozeau, J., Innes-Brown, H., & Blamey, P. J. (2013). THE EFFECT OF TIMBRE AND LOUDNESS ON MELODY SEGREGATION. *Music Perception, 30*(3), 259-274. doi: 10.1525/mP.2012.30.3.259

Marozeau, J., Innes-Brown, H., Grayden, D. B., Burkitt, A. N., & Blamey, P. J. (2010). The Effect of Visual Cues on Auditory Stream Segregation in Musicians and Non-Musicians. *Plos One, 5*(6). doi: 10.1371/journal.pone.0011297

McAdams, S., Winsberg, S., Donnadieu, S., & De Soete, G. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research/Psychologische Forschung, 58*(3), 177-192. doi: 10.1007/BF00419633

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2012). Introducing a new test battery and self-report inventory for measuring musical sophistication: The Goldsmiths Musical Sophistication Index. *Collected Work: Proceedings: The ICMPC--ESCOM 2012 Joint Conference—12th Biennial International Conference for Music Perception and Cognition; 8th Triennial Conference of the European Society for the*

Cognitive Sciences of Music. Pages: 705. (AN: 2012-19828).

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *Plos ONE, 9*(2), e89642.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport, 12*(1), 169-174. doi: 10.1097/00001756-200101220-00041

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin, 133*(5), 780-799. doi: 10.1037/0033-2909.133.5.780

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403-409. doi: 10.1016/j.tics.2009.06.003

Vliegen, J., Moore, B. C. J., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *Journal of the Acoustical Society of America, 106*(2), 938-945. doi: 10.1121/1.427140

Zendel, B. R., & Alain, C. (2009). Concurrent Sound Segregation Is Enhanced in Musicians. *Journal of Cognitive Neuroscience, 21*(8), 1488-1498. doi: 10.1162/jocn.2009.21140