The Effects of Familiarity on Musical Stream Segregation

Karen Stoskopf

Dr. Jessica Grahn

The Department of Physiology and Pharmacology, Schulich School of Medicine & Dentistry and The Department of Psychology, Brain and Mind Institute, The University of Western Ontario, London, ON, Canada

Abstract

The underlying mechanism for how the brain perceptually groups musical streams from a mixture to form coherent representations is unknown. Whether this mechanism requires prior experience with component streams in the mixture remains controversial. The current study compares two well-established mathematical models of stream segregation; matching pursuit, which requires prior experience with musical streams, similar to a knowledge-based process, and blind source separation, which does not require previous experience. The purpose of the study was to examine the role of familiarity in musical stream segregation to clarify which model best represents a potential biological mechanism. Undergraduate participants' (N =26, M_{age}= 19.42) ability to identify the number of non-lyrical musical streams in a mixture of either familiar or unfamiliar songs was measured. Familiarity was induced by exposing participants to the same 6 songs at the beginning of each trial block. Mixtures included two, three, or four overlapping songs. It was hypothesized that a greater familiarity with musical clips would increase participants' ability to separate the different streams in a mixture. We predicted that accuracy would increase with each successive trial block for familiar songs only and that greater accuracy would be observed for separating 2 overlapping songs than 3 or 4. Participants' accuracy was significantly greater for familiar clips compared to unfamiliar clips. Accuracy did not significantly improve across the trial blocks. Accuracy was significantly greater for 2 and 3 overlapping songs compared to 4 overlapping songs. Finally, familiarity significantly improved performance for 2 overlapping songs but not for 3 or 4 overlapping songs. This study suggests that musical stream segregation relies on familiarity, consistent with a knowledge-based process or matching pursuit-like model. Future studies focused on elucidating a potential biological mechanism should additionally consider the influence of experience and memory.

Keywords: blind source separation, knowledge-based (top-down) processing, matching pursuit, overlapping streams, stimulus-based (bottom-up) processing

Introduction

Perceiving music is a complex cognitive process (Platel *et al*., 1997). Musical pieces contain changes in pitch, intensity, rhythm and melody in both sequential and simultaneous dimensions (Bregman, 1990). Sequential dimensions include sounds that are connected over time and are perceived together, for example the melody of a song. Simultaneous dimensions include different acoustic properties that overlap over time but are from different sources, this would be equivalent to the harmony of different melodies produced by different instruments. The listener must therefore be able to perceive structural components, such as melody and harmony, as both separate and/or as a whole (Bregman, 1990). This challenges the listener with a rich set of perceptual possibilities, thus highlighting that perceiving music may require extensive cognitive processing (Pressnitzer *et al*., 2011). Despite these challenges, humans are equipped with an amazing ability to use specific auditory cues to separate and discern sounds, including musical sounds, from our environment. The mechanism by which the brain performs these tasks, especially in the context of music perception remains unknown. As a result of the complexity of music perception, it has become a popular research topic for understanding how the brain separates sound.

Our environment consists of many complex auditory signals from a variety of different sources. In order to understand and accurately represent our surroundings the mixture of sound reaching our ears must be correctly organized and grouped. The process by which the auditory system makes sense of our surroundings is called auditory scene analysis (ASA) (Bregman, 1990). The goal of ASA is to segregate different streams of sound that “go together” to make a coherent representation. A single stream can be defined as the sounds from a single source, and the separation of streams from a mixture is known as auditory stream segregation (Bregman, 1990). For example, the Cocktail Party Phenomenon (Cherry, 1953) is often used to describe perceiving speech in a noisy environment; stream segregation in this situation involves grouping aspects of a person talking beside you separately from the background noise of the room. Different cues are used by the auditory system to determine whether sounds are perceived as coherent streams or separate streams (Bregman and Campbell, 1971). The concept of coherence and streaming has been heavily researched in its application to music. Composers can take advantage of the mechanisms of stream segregation to promote illusions that facilitate the emergence of one or more melodic lines or fuse melodic lines so they are heard as one. The perception of music not only relies heavily of the process of ASA but also must defeat stream segregation properties to perceive the whole musical piece as coherent; fusion and segregation must be carefully controlled (Bregman, 1990). In addition, not only do these principles of stream segregation apply to perceiving individual songs but also apply to separating multiple songs playing at once. Despite advances in auditory research, very little is understood regarding individual differences in perceptual capabilities for stream segregation, especially its application to music and underlying mechanism (Bregman, 1990).

Although the mechanism is unknown, the enjoyment of music is universal has been an important part of life throughout history. To enjoy music and perceive its complexities the brain must separate streams via musical stream segregation (MSS). MSS more specifically refers to the application of stream segregation principles to musical streams. A musical stream refers to a perceived melodic line and depending on the acoustic properties of the stream, a musical stream can be a single melodic line of a song or the entire song. For the purpose of this study we define a musical stream as a single song. By understanding how healthy humans process music stimuli one can apply knowledge from such research to the treatment or therapy of pathological conditions, such as music therapy for Parkinson's disease or improving music perception in cochlear implants (Drennan and Rubinstein, 2008; Pacchetti et al., 2000). The objective of this study was therefore to gain a better understanding of a potential mechanism for MSS by looking at how humans separate multiple songs in a mixture.

Bregman (1990) proposed two different mechanisms for stream segregation, the first is a primitive or bottom-up process and the second is a knowledge-based or top-down process. Primitive processes are thought to be innate, meaning that stream segregation is primarily influenced by acoustic properties of the sounds and not by a higher form of knowledge, such as experience, memory or attention. Primitive processes have been implicated in the perception of speech (Darwin and Carlyon, 1995) and music (Bregman, 1990). In contrast, a knowledge-based process suggests that stream segregation involves experience, memory or attention, which have also been implicated in music perception. Dowling *et al*. (1973) demonstrated that universally familiar melodies such as Twinkle Twinkle Little Star were more readily identified when played at the same time as an unfamiliar distracter melody. Another study found that when participants heard a novel melody prior to hearing a mixture of melodies containing the same melody they just heard, song recognition for that melody was improved (Bey and McAdams, 2002) underscoring the role of experience in MSS. Some controversy still exists regarding which mechanism, primitive or knowledge-based, is most appropriate for musical stream segregation. Previously established computational models relevant to each of these processes can be used to study each mechanism.

Computational auditory scene analysis (CASA) uses computational means to study ASA (Bregman, 1990). CASA uses computer algorithms to model the way the human auditory system separates sounds. The complexity of the human auditory system has made finding computational models of auditory processing a challenging task (Haykin and Chen, 2005; Makeig *et al*., 1997; Stone, 2004). Two well-established computational models that can be used to study MSS are blind source separation (BSS) and matching pursuit (MP). These two models represent the opposite ends of a spectrum of potential mechanisms for MSS, with no prior experience with musical streams at one end and prior experience with musical streams at the other. BSS involves the separation of signals from a mixture with no prior information about the component signals, similar to a primitive process, while MP separates the signals using known information, similar to the knowledge-based process of stream segregation.

BSS is achieved by independent component analysis (ICA), using statistical similarities to group components together and to extract these signals from the mixture separately (Stone, 2004). ICA is based on the assumption that the source signals are statistically independent from each other. This means stream A provides no information about stream B and assumes that these independent signals are from different physical events (Stone, 2004). Therefore BSS would use statistical analyses of incoming sound to separate the incoming signals on the basis of independence. Some studies have shown that different neurons might encode different physical attributes; specifically this has been demonstrated in primary visual neurons (Barlow, 1989; Makeig *et al*., 1997). In this case, it was argued that BSS does provide a powerful method for identifying these independent sounds and relaying information to the primary sensory areas. Although BSS may be a good model for auditory processing of simple signals that are unfamiliar to the listener, some studies have argued BSS is an unrealistic assumption of the neurobiological context (Haykin and Chen, 2005). If BSS alone accurately portrays MSS, then previous exposure with a set of auditory stimuli would not be required to improve recognizability, however, much of human perception, both visual and auditory, relies on sensory experience and memory (Pinel, 2007).

In contrast, MP incorporates aspects of previous experience with streams to aid recognition of the sounds. MP uses a dictionary of known signals to decompose a mixture into its components. In other words MP uses a "best matching" technique to match incoming signals with known signals in a fixed dictionary to determine information about the individual signals in a mixture. For example, incoming signals from a flute will be compared to sounds of different instruments in the MP dictionary in order to identify a flute as the source of the incoming signals. MP models have been proposed that are capable of effectively separating sounds originating from different sources and different frequency ranges from background noise (Chu *et al*., 2008). A study by Peretz *et al*. (2009) suggests that recognition of a familiar tune in humans requires a selection procedure involving a memory system of all musical information from a lifetime, termed a musical lexicon. This musical lexicon is similar to the dictionary of signals used by MP. To investigate how similar the theoretical model of MP is to the physiological mechanism in humans, one needs to test the influence of familiarity on separating streams. Evidence that participants' ability to separate musical streams is enhanced if the musical streams are familiar supports arguments that a MP-like mechanisms is biologically plausible, in preference to BSS.

A number of studies suggest the existence of a constant dialogue between the auditory cortex and other cortical and sub-cortical brain regions indicating the importance of higher-level neural processes in music perception, such as memory, emotion or attention (Bigand *et al*., 2000; Dowling *et al*., 1987; Janata *et al*., 2002). Familiar melodies likely have stronger mental representations than novel melodies and, therefore, should be easier for the brain to segregate if memory is involved in the mechanism (Snyder and Alain, 2007). Studies have shown that talker familiarity can influence the ability to separate different streams of speech (Barker and Newman, 2004; Newman and Evers, 2007). Familiar voices were identified from the background more readily than unfamiliar voices (Newman and Evers, 2007). To investigate whether or not familiarity has an effect on MSS, participants' ability to identify the number of musical streams present in a mixture of two, three or four overlapping songs was measured. Participants underwent short-term familiarization by listening to six unfamiliar songs (so that they became familiar) and later identified how many streams of music were heard in a series of mixtures of either overlapping familiar (from familiarization) or unfamiliar songs. It was predicted that since many faucets of auditory perception rely on experience and memory (Pinel, 2007), MSS would as well, consistent with a knowledge-based or MP model of segregation. Therefore, it was hypothesized that greater familiarity with musical clips would increase participants' ability to separate them from a mixture. It was also predicted that accuracy would increase with each successive trial block for familiar songs only. Lastly, the effects of number of streams was examined, and it was predicted that accuracy for a simpler task of two overlapping songs would be greater than compared to a more difficult task of separating three or four overlapping songs.

Methods

Participants

Twenty-six undergraduate students were recruited from the University of Western Ontario Psychology Research Participation Pool. Each participant was compensated 1.5 research credits for just over an hour of participation in the study. Half of the participants were randomly assigned to group A (N=13) and received set A stimuli and half of the participants to group B

(N=13) and received set B stimuli (see procedure section). The study was approved by the Ethics Board of the University of Western Ontario and all participants gave informed consent.

Stimuli

Participants were presented with two different auditory stimuli. To begin each trial block, participants were presented with familiarization stimuli during the familiarization phase and later were presented with testing clip stimuli during the testing phase of the study.

All 62 songs used for the stimuli were uploaded from the Grahn Lab Music Database consisting of a collection of unfamiliar non-lyric music. Songs were categorized in terms of their genre, instruments and intensity. Genres included classical piano, classical violin, classical other, jazz, soft rock, rock, electronic or alternative. Intensity was subcategorized into four groups, very soft, soft, medium and loud.

Familiarization stimuli

Participants heard the first thirty seconds of six different songs, played one at a time, during the familiarization phase of the study. The familiarization songs were chosen in terms of their intensity; all six songs from the medium intensity subcategory. Genres and instruments of the six songs chosen varied. Of the six songs three were jazz (with different instrument combinations), one classical violin, one classical piano, one soft rock.

Testing clips stimuli

Participants heard twelve testing clips per trial block. Stimuli were composed of clips of a mixture of songs playing simultaneously for ten seconds. Mixtures were composed of two, three or four overlapping songs. All songs in the mixture were combined using Audacity software and volumes were normalized (DC offset set to one dB) and exported as wav files.

Overlapping songs were chosen primarily in terms of intensity subcategories and secondly in terms of similar genres and instruments. For example louder punk rock music would be overlapped with louder rock music or softer piano music overlapped with soft violin music. Mixing intensities proved to be difficult, as louder songs overpowered softer songs and normalization did not prevent this. Testing clips were either familiar, meaning that participants had previously heard each of the songs in the mixture in the familiarization phase, or testing clips were unfamiliar and comprised of novel overlapping songs.

Materials

The experiment was programmed and run using E-Prime (2.0) software in which the order of the trail block presentation as well as the order of stimulus presentation within the familiarization and testing phases were randomized. After the experiment participants completed a survey of their rhythm ability, which was used as a measurement of their musical experience. The survey included questions about music exposure, formal musical and dance training, if participants had any prior experience they were able to expand on the extent of the experience. Lastly the survey provided participants with a chance to state any strategies used in the experiment and leave any comments to the experimenter. The experiment and survey were presented on a Dell Precision laptop computer with Bose Acoustic Noise Cancelling headphones.

Procedure

Practice testing clips were provided prior to commencing the experiment so that the participants could become accustomed to hearing the testing stimuli and using the computer keypad to make responses. Participants were presented with four practice testing clips, all clips were overlapped using the same procedure as the testing stimuli however they were composed of songs that were not used in the experiment. Two of the practice testing clips contained two overlapping songs and two of the practice testing clips contained three overlapping songs. Responses from the practice trials were not recorded.

The experiment contained three trial blocks, each with a familiarization phase followed by a testing phase. Participants began each trial block by listening to the first thirty seconds of six different songs (familiarization phase) followed by the testing phase. During the testing phase participants listened to testing clips and responded, using the computer keypad, by indicating how many streams of music they heard in the mixture. Testing clips were either composed of ten seconds of the thirty seconds of the familiar songs previously heard in the familiarization phase or ten seconds of novel songs. A total of twelve testing clips (six familiar and six unfamiliar) were presented per trial block. Of the six familiar clips, two of the clips were a mixture of two overlapping songs, two of the clips were three overlapping songs and two of the clips were four overlapping songs. The same breakdown of unfamiliar clips was used; two of the unfamiliar clips were a mixture of two overlapping songs, two of the clips were three overlapping songs and two of the clips were four overlapping songs. The second and third trial block consisted of the same familiarization stimuli in order to induce a short-term increase in familiarity with those six stimuli across time. Participants in both group A and B each received familiarization and testing clip stimuli however the songs used in each group were different in order to determine the replicability of the results. Group A participants were presented a set of six different familiarization songs than Group B participants. After the experiment participants completed the survey of musical ability and experience and were debriefed.

Design

Participants' accuracy, the proportion of trials in which the number of songs was correctly identified (proportion correct) was measured. Next, a 2 [Familiarity (unfamiliar or familiar songs)] X 3 [(Number of overlapping songs; 2,3 or 4)] X 3 [(Time point; trial block 1,2 and 3)] repeated measures ANOVA was conducted with a Bonferroni post-hoc test.

Results

Data from 26 participants (M_{age}= 19.42, *SD* = 1.58) were exported from E-Prime (.edat file) and into SPSS for analysis. The proportions of correct trials of each condition were averaged for each participant and mean accuracy across all participants was calculated. Repeated measures ANOVA and Bonferroni post-hoc test revealed a main effect of familiarity (F (1, 26) = 20.57, $p < 0.001$) indicating that participants' mean accuracy was significantly greater for familiar clips (M = 0.515) than unfamiliar clips (M = 0.370), as predicted (Fig. 1). A main effect of the number of overlapping songs (streams) was also found (F (2, 26) = 29.18, $p < 0.001$) (Fig. 2). Participants' accuracy for two overlapping songs (M = 0.561) was not significantly different than for three overlapping songs (M = 0.494), but was significantly greater than for four overlapping songs (M = 0.272; F (1,26)= 44.14, $p < 0.001$). Further, accuracy for three overlapping songs was significantly greater than four overlapping songs (F (1,26) = 33.20, $p < 0.001$). Figure 2 showed a significant decrease in accuracy with greater numbers of overlapping songs. Two overlapping songs had the highest accuracy and four overlapping songs had the lowest accuracy. In contrast to our hypothesis, no significant differences were found between accuracies across the time points indicating that performance did not improve over the experiment (Fig. 3). Although non significant, a trend of increasing accuracy over time was observed, with accuracy lowest for the first trial block and highest for the third trial block (M_1 = 0.404, M_2 = 0.458, M_3 = 0.465) (Fig. 3). Contrary to our hypothesis, no interaction between familiarity and time point was observed and thus no differences between accuracy of familiar and unfamiliar conditions across time. A familiarity X number of overlapping songs interaction was found (F (2, 26) = 4.44, $p < 0.05$) indicating that differences in the accuracies for the different number of streams was affected by familiarity (Fig. 4). Significantly improved performance for two overlapping songs (M = 0.692) compared to three overlapping songs (M = 0.538) was observed in the familiar condition only (F (1, 26) = 6.43, $p < 0.05$). Similarly, significant differences were observed between familiar and unfamiliar conditions in performance for two overlapping (familiar M = 0.692, unfamiliar M = 0.429) compared four overlapping songs (familiar M = 0.314, unfamiliar M = 0.231; F (1, 26) = 5.47, $p < 0.05$). No significant differences were observed across familiarity conditions for accuracy of three compared to four overlapping songs. Thus, familiarity significantly improved performance for two overlapping songs but not three or four overlapping songs.

Discussion

The purpose of the present study was to examine the role of familiarity in MSS to clarify which computational model, either BSS or MP, best represents a potential biological mechanisms. Consistent with our hypothesis, familiarity significantly improved musical stream segregation supporting a MP-like model. This suggests that familiarity and experience affect the processing of musical streams. Similar to speech processing, familiar voices or musical streams in a mixture may lead to different processing or enhanced perception (Newman and Evers, 2007). Previous studies demonstrate the role of cortical brain regions (which process higher forms of knowledge) in processing musical sounds; Platel *et al*. (1997) found that familiarity with musical streams lead to a greater activation in the anterior parts of the left hemisphere similar to speech, consistent with Peretz *et al*. (2009) who suggest that a musical lexicon (much like a phonological lexicon used in speech processing) exists in the superior temporal sulcus. The use of best-fit matching of musical input with prior information about musical streams from a musical lexicon parallels a MP model for stream segregation. Both neuroimaging studies provide strong evidence for enhanced processing of familiar songs by a musical memory system. Future neuroimagining studies focusing on whether or not similar activation is observed for familiar stream segregation would help bridge the gap between familiarity research and MSS.

Furthermore, we predicted that accuracy would be greater for two overlapping songs, a simpler task, compared to three or four overlapping songs, a more difficult task. Consistent with previous research, participants were able to discern overlapping melodies (Bey and McAdams, 2002; Dowling, 1973; Hartmann and Johnson, 1991) and accuracy declined with increasing number of overlapping songs. One explanation is that with an increase in number if overlapping songs, the time to focus/identify each song in the mixture decreases. For two overlapping songs, participants could use half of the ten seconds to attend to one stream and half for the other, in contrast to a third or a fourth of the time for three and four overlapping songs. This is in line with familiarity becoming poorer with faster presentation demonstrated by Bregman (1990). Unexpectedly, the effect of familiarity differed for two overlapping songs compared to three and four overlapping songs. The familiarity by number of overlapping songs interaction revealed that familiarity significantly improved performance for two overlapping songs but not for three or four. The simpler task of two overlapping songs may have provided enough time for the brain to access melodic representations in short-term memory. It is possible that the presentation of the clips was too brief for the brain to accurately process three or four overlapping songs and therefore there were no differences between the familiar and unfamiliar condition for three and four overlapping songs. Alternatively, many studies have demonstrated the problem of divided attention in stream segregation (Bregman, 1990; Dowling et al., 1973). Dowling *et al*. (1973) showed that participants perceived overlapping familiar tunes as separate streams, however, they had difficulty identifying both of them. Similarly, humans may not be able to pay attention to all steams at the same time, thus making more overlapping songs a more challenging task (Bregman, 1990). It is possible that in the present study participants were able to unconsciously perceptually group four overlapping streams but were unable to explicitly identify that there were 4 streams in the mixture. Although the significant decline in accuracy suggests that the more overlapping songs the more difficult the task, recognition of melodies in a mixture depends on several factors such as relatedness, timbre, pitch and tempo making it difficult to determine what factors are responsible for the decline in accuracy (Bregman, 1990; Gregory, 1990).

No main effect of time or time interaction was observed, contrary to our hypothesis. Accuracy did not significantly improve with each subsequent trial block. Even though results were non significant, a trend of increasing accuracy was observed. Perhaps the familiarization phase was not long enough to yield a significant increase in familiarity with each trial block. Another explanation is that the first trial block was enough to induce short–term familiarity and subsequent trial blocks served only to maintain the same level of familiarity, and the small change in accuracy over time was due to learning and practice with the task itself. Familiarity with musical streams in real life may be different from short-term familiarity induced in a laboratory setting. Different degrees of familiarity are not well distinguished in the literature (Newman and Evers, 2007). For example, there are stricking differences between the level of familiarity in the following studies: the universally familiar melodies such as Twinkle Twinkle Little Star used in the Dowling *et al*. (1973) compared to an increased exposure to songs with each trial block in the present study or the single presentation of a novel song prior to hearing a mixture containing that same novel song in the Bey and McAdams (2002) study. To fully determine the role of familiarity, future studies should consider the effects varying degrees of familiarity on participants' stream segregation ability.

Similarly, degree of familiarity can influence how the streams are perceived and interpreted. Explicit knowledge of streams helps listeners attend to that particular stream and familiarity helps interpret speech appropriately (Newman and Evers, 2007). This finding emphasizes the influence of explicit vs. implicit familiarity on perception. A limitation of the present study is that we did not ask participants "what streams do you hear in the mixture." It is therefore unclear whether the significant increase in accuracy for familiar clips resulted from an explicit recognition of songs from the familiarization phase or an implicit recognition. It is possible this could influence how and what stored perceptual representations are activated.

Another limitation of the present study is the complexity and ambiguity of music stimuli. Often jazz music promotes illusory fusion of streams, blending sources so that they appear to be coherent (Pressnitzer *et al*., 2011). The use of jazz songs for three out of six of the familiarization stimuli may therefore have promoted fusion of streams. In the case of fusion participants may have underestimated the number of streams in the mixture. Music can also be heard as pseudo-polyphony, when a single stream is perceived as many (Pressnitzer *et al*., 2011) and in this case participants may overestimate the number of streams in the mixture. The ambiguity and complexity of music may explain individual variability in the accuracy for MSS. For this study any deviation from the correct response was considered an incorrect response, however, potential errors might have stemmed from overestimation or underestimation rather than inaccurate stream segregation. The cause of incorrect responses should therefore be investigated to determine if certain mixtures used in this study were promoting fusion or pseudo-polyphony and whether individual differences exist. Further, musical training could affect how participants perceive fusion or pseudo-polyphony. Due to the musical nature of the study, almost all participants were musically experienced and only four participants had no formal musical training. Familiarization with an instrument or genre of music could increase participants' sensitivity to musical properties such as pitch changes thereby improving MSS ability. Future studies should investigate the effects of musical training on MSS.

In conclusion, familiarity plays a role in MSS suggesting that experience and short-term memory can influence stream segregation, consistent with a knowledge-based process or MP-like model. This however does not completely discredit the potential for a combination of the two mechanisms working together. Knowledge-based segregation may interact with lower-level perceptual processes like a primitive process to accurately achieve MSS. Although this study supports a MP-like model further research needs to be conducted to fully elucidate a potential biological mechanism for MSS. Future research on music perception will not only help with further understanding how a healthy human brain separates simultaneous streams but can also help in diagnosing stream segregation issues experienced in hearing disorders and, therefore, maximize potential treatments. Applications of such research can help improve otherwise poor music perception in cochlear implants which can have a very positive impact on cochlear implantees' lives (Drennan and Rubinstein, 2008).

Acknowledgements

A special thanks is given to Dr. Jessica Grahn for her continued support and encouragement throughout the process of my Honours thesis. This project would not have been possible without her guidance and expertise. I would also like to thank Dr. Mark Daley as well as the Grahn lab for all their knowledge and advice.

Literature Cited

Barker BA, Newman RS. 2004. Listen to your mother! The role of talker familiarity in infant streaming. Cognition 94:B45-B53.

Barlow HB. 1989. Unsupervised learning. Neural Comput 1:295-311.

Bey C, McAdams S. 2002. Schema-based processing in auditory scene analysis. Attention, Perception, & Psychophysics 64:844-854.

Bigand E, McAdams S, Foret S. 2000. Divided attention in music. International Journal of Psychology 35:270-278.

Cambouropoulos E. 2008. Voice and stream: Perceptual and computational modeling of voice separation. Music Perception 26:75-94.

Carlyon RP. 2004. How the brain separates sounds. Trends Cogn Sci (Regul Ed) 8:465-471.

Cherry EC. 1953. Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25:975-979.

Chu S, Narayanan S, Jay Kuo CC. 2008. Environmental sound recognition using MP-based features:1-4.

Darwin C, Carlyon RP. 1995. Auditory grouping. The Handbook of Perception and Cognition 6:387-424.

Dowling WJ. 1973. The perception of interleaved melodies. Cognit Psychol 5:322-337.

Drennan WR, Rubinstein JT. 2008. Music perception in cochlear implant users and its relationship with psychophysical capabilities. Journal of Rehabilitation Research and Development 45:779.

Hartmann WM, Johnson D. 1991. Stream segregation and peripheral channeling. Music Perception:155-183.

Haykin S, Chen Z. 2005. The cocktail party problem. Neural Comput 17:1875-1902.

Janata P, Birk JL, Van Horn JD, Leman M, Tillmann B, Bharucha JJ. 2002. The cortical topography of tonal structures underlying Western music. Science 298:2167-2170.

Makeig S, Jung TP, Bell AJ, Ghahremani D, Sejnowski TJ. 1997. Blind separation of auditory event-related brain responses into independent components. Proceedings of the National Academy of Sciences 94:10979-10984.

Newman RS, Evers S. 2007. The effect of talker familiarity on stream segregation. Journal of Phonetics 35:85-103.

Peretz I, Gosselin N, Belin P, Zatorre RJ, Plailly J, Tillmann B. 2009. Music Lexical Networks. Ann N Y Acad Sci 1169:256-265.

Pinel JP. 2007. Biopsychology:Pearson Education

Platel H, Price C, Baron JC, Wise R, Lambert J, Frackowiak R, Lechevalier B, Eustache F. 1997. The structural components of music perception. A functional anatomical study. Brain 120:229-243.

Pressnitzer D, Suied C, Shamma SA. 2011. Auditory scene analysis: the sweet music of ambiguity. Frontiers in Human Neuroscience 5

S B, Albert. 1990. Auditory scene analysis :the perceptual organization of sound. Cambridge, Mass.:MIT Press. 773 p.

Snyder JS, Alain C. 2007. Toward a neurophysiological theory of auditory stream segregation. Psychol Bull 133:780.

Stone JV. 2002. Independent component analysis: an introduction. Trends Cogn Sci (Regul Ed) 6:59-64.

Appendix

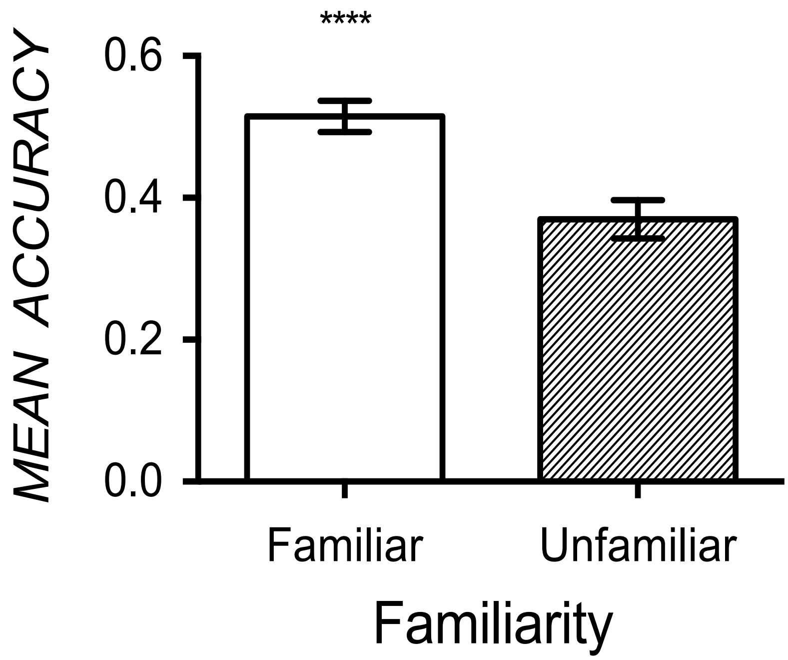

Figure 1. Effects of familiarity on the mean accuracy for separating overlapping songs (combination of accuracy for 2,3 and 4 overlapping songs). Comparisons between familiar and unfamiliar clips demonstrate a significantly greater accuracy in the familiar condition ($p < 0.001$). Data shown are the mean accuracy ± standard error of the mean, N=26.

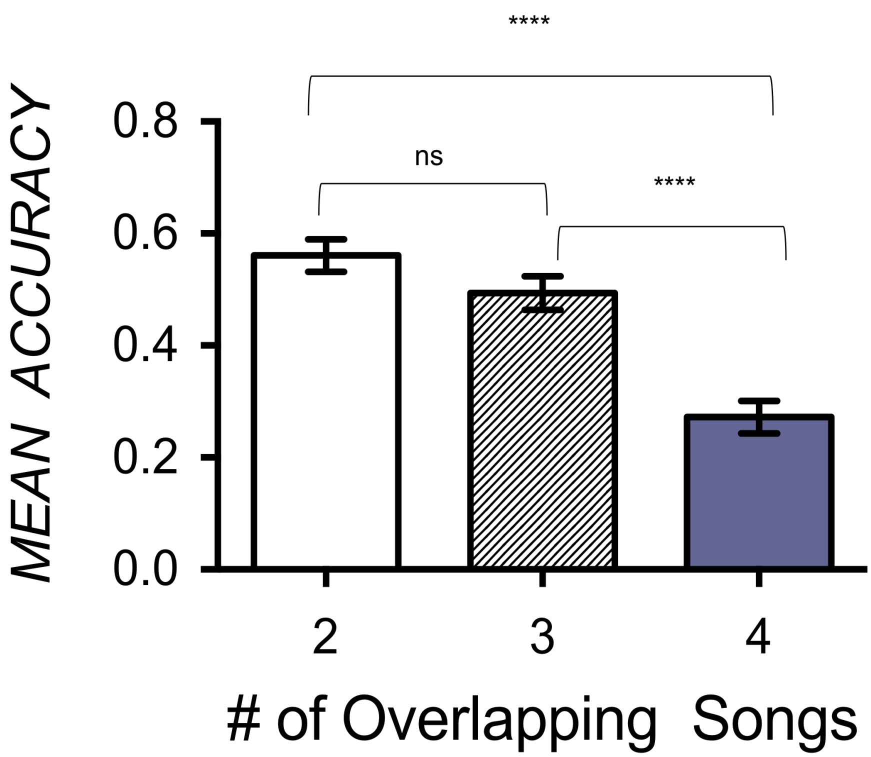

Figure 2. **The effect of the number of overlapping songs on participants mean accuracy from combined familiarity conditions (familiar and unfamiliar).** No significant differences in participants' ability to separate two overlapping songs compared to three overlapping songs were observed (p > 0.05). Mean accuracies were significantly greater for two overlapping songs in comparison to three overlapping songs ($p < 0.001$) as well as significantly greater for three overlapping songs in comparison to four overlapping songs ($p < 0.001$). Data shown are the mean accuracy ± standard error of the mean, N=26.

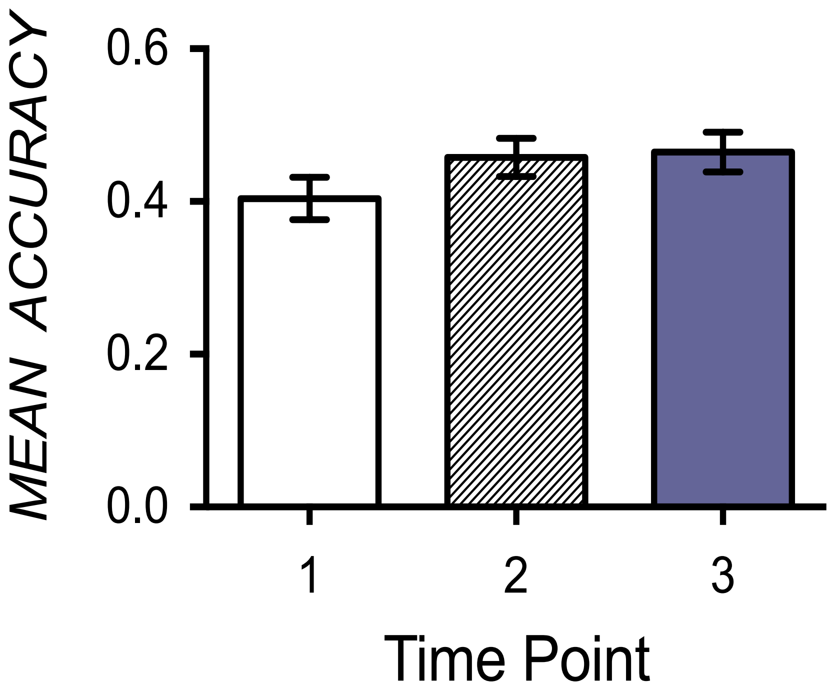

Figure 3. Mean accuracies across time. Time point represents the order in which the trial blocks were presented, 1 being the first and 3 being the last. No significant differences observed ($p > 0.05$). However, an observed trend of increasing accuracy with each successive trial block; the first trial block with the lowest accuracy and the third trial block with the highest accuracy. Data shown are the mean accuracy ± standard error of the mean, N=26.

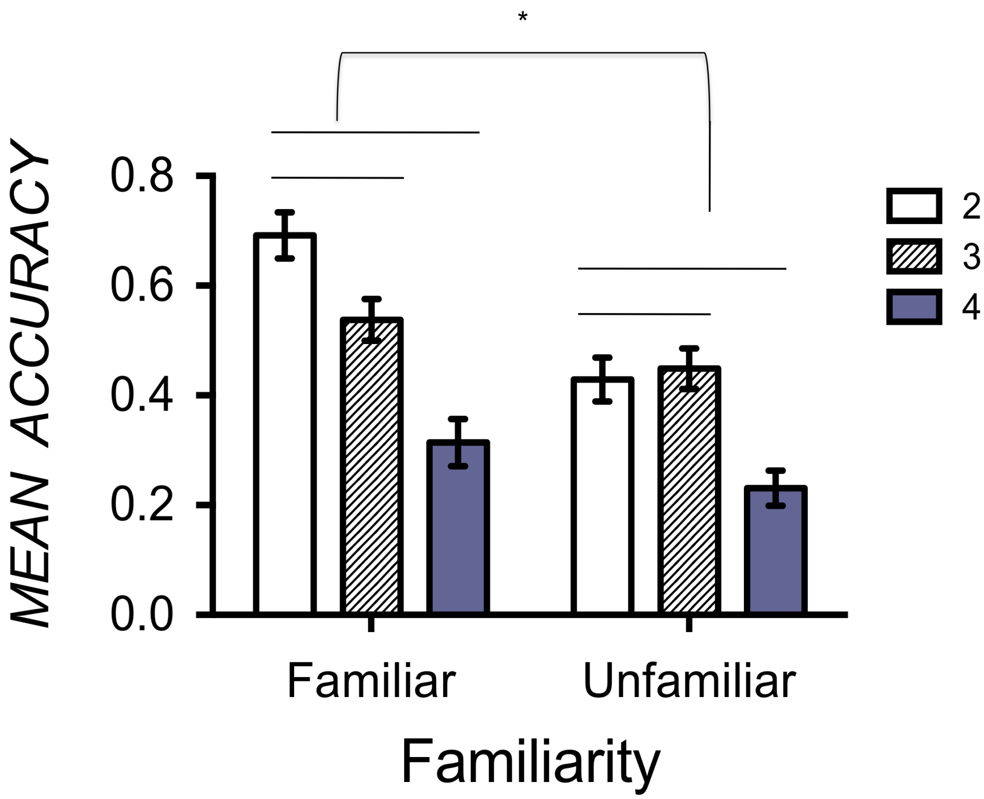

Figure 4. Effects of familiarity condition on participants mean accuracy for 2,3 and 4 overlapping songs. Mean accuracies for separating overlapping songs shown for each familiarity condition (familiar and unfamiliar). The legend denotes the number of overlapping songs; 2: two overlapping streams, 3: three overlapping streams and 4: four overlapping streams. The main effect of overlapping songs was affected by familiarity therefore significant differences were observed between familiar and unfamiliar conditions ($p < 0.05$). Significant differences in two compared to three overlapping songs as well as two compared to four overlapping songs between the familiar and unfamiliar condition ($p < 0.05$). No significant differences between familiar and unfamiliar condition for three compared to four overlapping songs ($p > 0.05$), indicating the familiar condition significantly improved performance for two overlapping songs but not three or four overlapping songs. Data shown are the mean accuracy ± standard error of the mean, N=26.