NEURAL CORRELATES OF ATTENTION IN AUDITORY STREAM SEGREGATION

Anahita Mehta

The Ear Institute

Faculty of Brain Sciences

University College London

Submitted in partial fulfilment for the degree of Doctor of Philosophy

August 2015

DECLARATION

This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

In accordance with the University College London guidelines, this thesis does not exceed 100,000 words, and it contains less than 150 figures.

Signed:………………………………………………………………………………

Date:…………………………………………………………………………………

Anahita Mehta

London

SUMMARY

The work in this thesis illustrates how directed attention can modulate multistable, ambiguous auditory percepts, and how these percepts are reflected in the stimulus-driven cortical EEG responses of human listeners. Natural auditory environments require listeners to parse out an acoustic signal of interest amidst auditory sources constantly overlapping or competing for salience. Listeners need to simultaneously use both sequential and synchronous sound segregation to focus on a target as most sounds both overlap and unfold over time. Additionally, understanding how ambiguous stimuli are perceived as well as represented in brain activity can be useful in dissociating the neural responses to physical stimuli from the correlates of perception.

The first set of experiments explored the effect of attention on sequential sound segregation using a perceptually ambiguous stimulus described by van Noorden (1975). Following on from these findings, a novel stimulus based on a variant of Deutsch's 'octave illusion' (Deutsch, 1974), which involved ambiguous stimuli that engaged both synchronous and sequential sound segregation, was investigated. The experiments using this new stimulus paradigm demonstrated that the octave illusion was subserved by the same mechanisms that govern auditory streaming. Furthermore, directed attention could alter the percept of this stimulus and these changes could be observed in the corresponding cortical brain activity. Subsequent experiments were carried out to further understand the mechanisms underlying the octave illusion. Results from psychophysics, cortical EEG and modeling consistently suggested that the perceived illusory percept results from a misattribution of time across perceptual streams of synchronous sounds.

Overall, the results highlight the key role of attention in complex auditory stream segregation involving both alternating as well as synchronous sound segregation. This body of work also introduces a stimulus, typically associated with an auditory illusion that has not previously been studied with performance-based behavioral measures as a versatile and experimentally valuable stimulus to study stream segregation.

ACKNOWLEDGEMENTS

This PhD has been an incredible experience and the following list of people to thank violates graduate school rules relating to brevity.

I would like to immensely thank Ifat Yasin for her infinite patience with me, for providing me with the freedom and constant support to pursue whatever I wanted and encouraging me throughout this process. I have been very fortunate to have a supervisor as caring and dedicated as her.

I am indebted to Shihab Shamma for his inexhaustible encouragement, unconditional support and most importantly, for making this scientific journey such fun! His incredible energy, extreme hard work and inimitable ability of considering his students as peers are qualities that I hope to emulate.

I am equally grateful to Maria Chait for being an incredible role model and mentor and for always looking out for me. The dedication and care that she demonstrates towards her work, family and students is inspiring and I consider myself fortunate to have learnt from her.

I am also grateful to Andrew Oxenham for his invaluable time, suggestions and the inherent sense of clarity he brought to every scientific discussion. I am also very thankful to Andrew Faulkner for his constant support and care throughout these years. I would also like to appreciate the enthusiasm and incredible amount of hard work put in by Nori Jacoby into developing the new model.

I would also like to thank David McAlpine and Debi Vickers for their valuable advice, Vit Drga, Jonathan Simon, Alain de Cheveigné and Jonathan Fritz for several thought-provoking discussions and advice on analysis, Anthony Chan for the technical support in Maryland, and Hedwig Gockel and Bob Carlyon for introducing me to the exciting world of auditory research.

This PhD would have been impossible to get through without my friends and family. Big thank you goes out to Nic (Bababa) (for being an incomparable PhD buddy and sharing hilarious times, dramatic moments, dishoom overdoses, skype chats, code frustration, obsession with poster aesthetics, and many miles of travel), Tejoo (for being a virtual room buddy, SOS friend, and fellow postcard junkie), Lara (for all the love, Hampstead adventures, late night experiments and introducing me to vegetables), Anu (for her unadulterated positive energy and love), Sundeep (for all the advice, tea and pesto), Larissa (for all the crazy laughs and bear hugs), Francisco (for all the books and lovely conversations), Lucie (for being a lovely host, participant and friend) and Shetty (for the bro codes).

A massive thank you to my Nottingham family for being an integral part of my life and for all the inside jokes. Thank you to Outi (for the public hugs, Shrek, coffee with fake names and ridiculous PhD stories), Helen (for wave V, GG, rustic coding, special animal lols and traditional but rare moments), Pete (for all the MATLAB, cooking, games and steve), Kate (for the inappropriate jokes, awkward foot-in-mouth moments, all things green and her over politeness stories) and Nick (for unbelievable laughs, great advice and wizard costumes).

I have been lucky to have Salma (for the running, disapproval of food habits and incredible cinema stories) and Pavi (for shelter, chilli oil, life dilemmas and thesis motivation) as neighbours and I am thankful to them for making the years at GC full of moments ranging from highly intellectual conversations to extreme laughter fits over many mugs of tea and Sunday brunches.

I am also thankful to all the wonderful students at the Ear Institute: Warren (for the love-hate friendship), Stefano (for late night PhD room chats), Jaime (for all the coding), Ed (for all the coffee), Lucile (for the restaurant discussions), Kath (for the fun conversations) and Elena (for making my room her own).

I would like to thank all my friends back home who make it very hard to say goodbye every January. Big thank you to Anup (for being an exceptional friend across several time zones, thwacking my head during my frequent judgemental moments and making me realize that there is more to life than stressing over stereo cables), Kanchoo (for the unreserved love, warmest welcomes, street food overdoses and for being friends since age 10), Gautam (for gokul, team players and beer), (Havs (for the ridiculous skype calls and the equally serious life discussions), Rinks (for the public displays of affection and always being up for travel plans), Naash and Ro (for all the drives, travel talk and bad jokes), Tini (for being the best party buddy), Zi (for all the cheer and support) and Di (for being there always).

Big thank you to my canary wharf friends; Pravs (for all the ridiculous Bollywood gossip, GG and sambal), Nezoo (for all the breakfast dates) and Laks (for nocturnal story sessions and being a lady of leisure) who saw me through the crazy first year of graduate school. I will always cherish all our trysts with experimental baking, python modes, excessive sitcoms and instant noodles.

Special thank you also goes out to Dimi, Alex, Francisco, Kotaro, Lakshmi, Varun, Dana, Brendon, Ifte, Krishna, Francisco and Nicholas for all the fun times in Maryland and Telluride.

A big thank-you especially to Kaashu for eating everything I cook and helping us stay cool and Koko for all the weekend chatter and cake.

Finally, this is for my closest four: Mummy (Popu) and Papa (for their unconditional love and support across continents, numerous early morning wake-up phone calls, overfeeding at airports, and for always being immensely proud of every achievement, however big or small) and Fred and Gudz (for believing in me from the start, going out of their way to take care of me, improving my food habits, reverse engineering food and never giving up on me).

CONTENTS

List of Figures

List of Abbreviations and Acronyms

AC	auditory cortex
AM	amplitude modulation
ANOVA	analysis of variance
dB	decibel
EEG	electroencephalography
ERP	evoked response potential
FM	frequency modulation
fMRI	functional magnetic resonance imaging
Hz	Hertz
MEG	magnetoencephalography
MMN	mismatch negativity
SNR	signal-to-noise ratio
SSA	stimulus specific adaptation

Chapter 1
INTRODUCTION

1.1 Auditory Scene Analysis and Stream Segregation

Natural auditory environments are inherently noisy with various auditory sources constantly overlapping or competing for salience. This has led to the study of the problem of parsing out the acoustic signal of interest amidst the cacophony of sounds, commonly referred to as the 'cocktail party problem' (Cherry, 1953). The problem, although seemingly limited to humans from the terminology, exists for most species, as most non-human animals frequently need to segregate sounds of interest for survival and locating mates and offspring (Fishman & Steinschneider, 2010).

Conceptually, the cocktail party problem presents two closely related, yet distinct challenges; the first being the problem of physical sound segregation in an acoustic scene and the second, the matter of directing one's attention to the sound of interest whilst suppressing other concomitant auditory signals (Bregman, 1990). The first aspect of physical segregation of the sound sources based on their physical characteristic varies according to the spectral and temporal properties of the target sound as well as the competing sound sources (Bregman, 1990). The second aspect of selective attention was highlighted by Cherry (1953) and in the context of speech perception, relates to how listeners attend to one speech signal over another and listeners' ability to switch attention between speakers.

This problem of sound segregation has received significant attention since the 1970s when it was studied in detail by Bregman & Campbell (1971) who put forth the concept of 'auditory scene analysis'. It was suggested that in everyday life, listeners are constantly surrounded by multiple sound sources (each being a distinct acoustic event), which is often referred to as an auditory scene. Listeners frequently need to analyse these interfering complex sound patterns into separate 'auditory streams' in order to hear out the target sound source. This process of parsing the various sounds has been commonly referred to as 'auditory scene analysis' (Bregman, 1990; Bregman & Campbell, 1971). The basic premise of this theory of 'auditory scene analysis' is that there isn't an exclusive method, which can disentangle these various sounds from a composite whole (Bregman, 1990; Denham & Winkler, 2006). Therefore, it has been suggested that the auditory system engages several sound analysis 'heuristics' that are based on specific properties characteristic of naturally occurring sounds (Darwin & Carlyon, 1995; Moore & Gockel, 2002). For example, most of the principles of stream segregation were originally based on Gestalt principles (Koffka, 1935). An example of the role of Gestalt principles in streaming can be illustrated by how the notes of a scale played in succession by a particular instrument are grouped together based on the instrument's timbre as well as the principle of 'good continuation' of the pitch sequence (Bregman, 1990).

Bregman (1990) suggests that if these auditory events need to be perceived as distinct and different, there needs to be 'a level of mental description' where each individual event is attributed an individual mental representation. An auditory stream could be described as a percept of successive and/or simultaneous sound elements that are either perceived as a coherent whole and appearing to emanate from a single source or perceived as more than one sound sources, in which case the sounds deemed to be coming from different sources are allocated separate auditory streams (Moore & Gockel, 2012). These streams, typically, can be selectively attended to and followed individually or together over time (Shamma & Micheyl, 2010).

The phenomenon of streaming has been described in terms of two broad groups of auditory events: simultaneously and sequentially occurring auditory components in an auditory scene (Bregman, 1990). Bregman (1990) also proposed a two-stage model, which is divided into primitive- and schema-based processes. Primitive processes are based on innate capabilities of the individual, which apply for most sounds (van Noorden, 1975). Schema-based processes that involve the learning of certain predetermined rules for auditory processing, are dependent on previous experiences of the listener and may not be applicable for all auditory situations (Bey & McAdams, 2002). The difference between the two processes (primitive and schema-based) can be understood by the following simple example; neonates and infants can easily segregate rapidly occurring low and high frequency sequential tones (McAdams & Bertoncini, 1997; Winkler et al., 2003) making this a primitive innate process (i.e. not learnt). However, the ability of an orchestra conductor to segregate and follow a particular instrument in a large ensemble of instruments is a difficult task which can only be carried out as a result of learning schema-based processes (Denham & Winkler, 2006). Bregman's two-stage model suggests that in the first stage, the incoming acoustic signal is parsed into streams automatically through the primitive processes. However, in the second stage, competition between the various sound streams is resolved based on a number of factors including attention and previously learnt rules (for example, grouping by timbre as described in the case of the orchestra conductor (Denham & Winkler, 2006).

We begin by describing the acoustic factors that play a role in auditory stream segregation.

1.2 Properties of auditory stream segregation

1.2.1 Spectral and temporal properties of stream segregation

One of the earliest experiments carried out for studying sequential stream segregation was by Miller & Heise (1950). They carried out a systematic experiment on an effect that they had incidentally observed in a previous experiment (Miller, 1947). They had noted that a rapidly alternating pattern of high and low frequency tones tended to perceptually ‘break up into two melodies’. They measured listeners’ percept of this rapidly alternating sequence as a function of frequency separation between the low and high frequency tones. They found that for a small frequency separation, the alternations were perceived as a continuous high-low-high-low (in frequency) tonal percept. However, when the frequency separation was larger, the alternating tone pattern was perceived as two unrelated tone streams where the high tones and low tones segregate into two parallel auditory streams. They called the transition point between these two perceptual organisations as the ‘trill threshold’ for the listener.

Van Noorden (1975) studied the initial findings of the ‘trill threshold’ with sequences of tone triplets of an ABA_ arrangement (A and B refer to pure tones of low and high frequencies). This type of tone sequence could be perceived either as an integrated percept where the sequence sounds like a galloping pattern of tones emanating from a single sound source or as a segregated percept where the galloping percept was lost and the sequence was perceived as two separate monotonous streams (for example, a monotonous stream can be depicted as AAAA or BBB in this context); each stream composed of a single repeating tone of sound at different rates (Figure 1.1 shows a schematic representation of these ABA_ tone triplets, depicted by ‘L’ and ‘H’ for Low and High frequencies, which can either be perceived as a one-stream integrated percept or a two-stream segregated percept).

He further analysed this perceptual behaviour of human listeners in response to these repeating pure tone triplets (ABA_) and characterized perceptual boundaries that governed integration, segregation and bistability of the perceptual response (van Noorden, 1975). Figure 1.2 indicates these boundaries. The temporal coherence boundary (TCB) was defined as the limit beyond which listeners could never hear one integrated galloping auditory stream whilst the fission boundary (FB) was defined as the limit beyond which listeners could never hear two segregated sound streams (Figure 1.2 shows a schematic representation of the fission and temporal coherence boundaries). The area between the two curves represented the 'bistable percept region', which was the spectral-temporal region within which the percept elicited by the ABA_ tone triplets was bistable and could switch between that of an integrated or segregated percept. Van Noorden (1975) also highlighted the key roles of frequency separation and rate of tone presentation in determining the perceptual state of the listeners (either integrated or segregated). A segregated, two-stream percept was more likely to be perceived under conditions of larger frequency separations and faster rates of presentation. This means that the temporal coherence boundary was affected by both these factors. The effect of presentation rate, however, was negligible when trying to segregate the streams, i.e., on the fission boundary. This indicated that if the frequency difference between the tones in a given sequence were small, the sequence would continue to sound as an integrated percept despite changes in presentation rate (Figure 1.2).

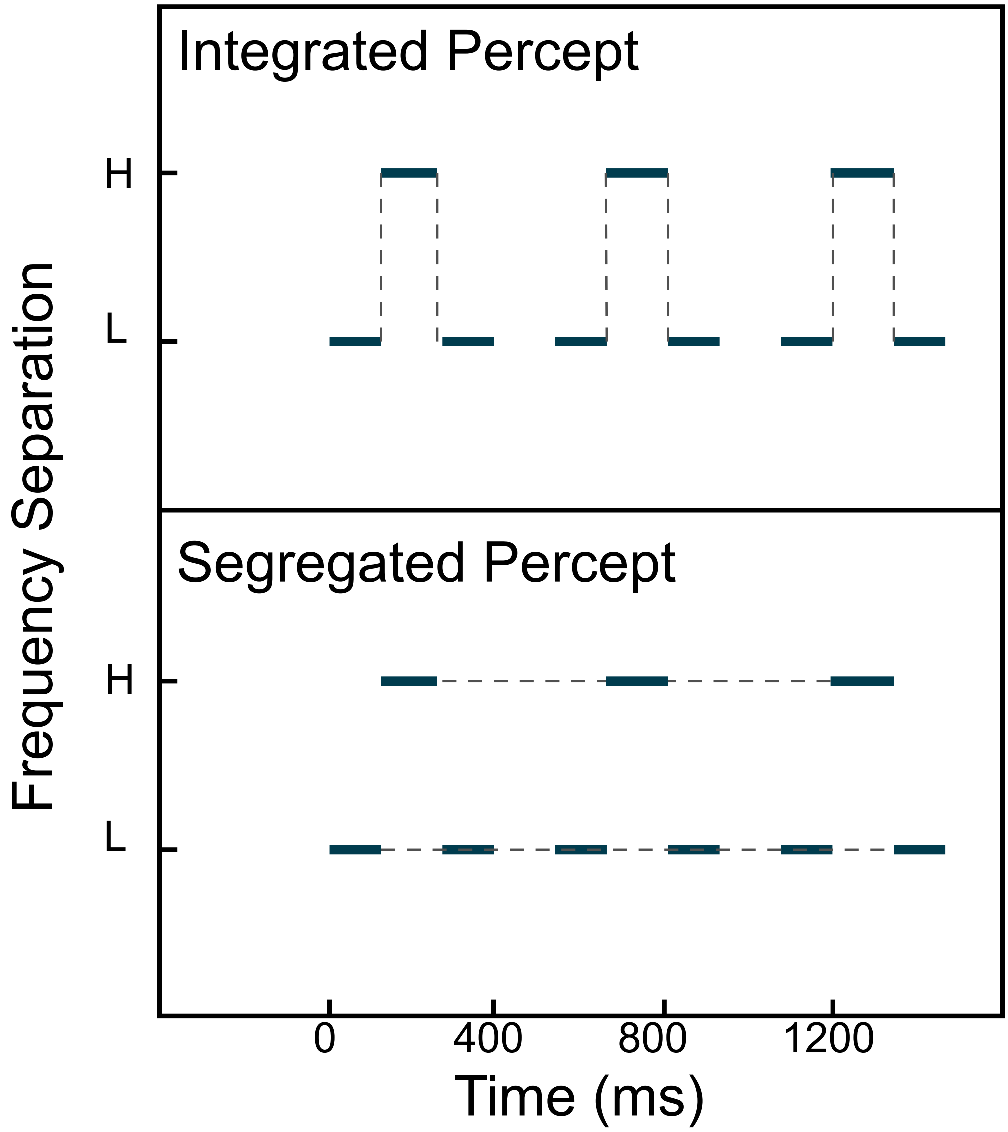


Figure 1.1: *Solid lines represent pure tones of either a low (L) or a high (H) frequency. The tones are arranged in an LHL-LHL-LHL- sequence (originally developed by van Noorden, 1975). Dashed grey lines represent perceptual grouping. In the top panel, the L & H tones are grouped together (integrated), whereas in the bottom panel, two separate streams are perceived (non-integrated), each corresponding to either the L or the H pitched tones. Adapted from van Noorden (1975).*

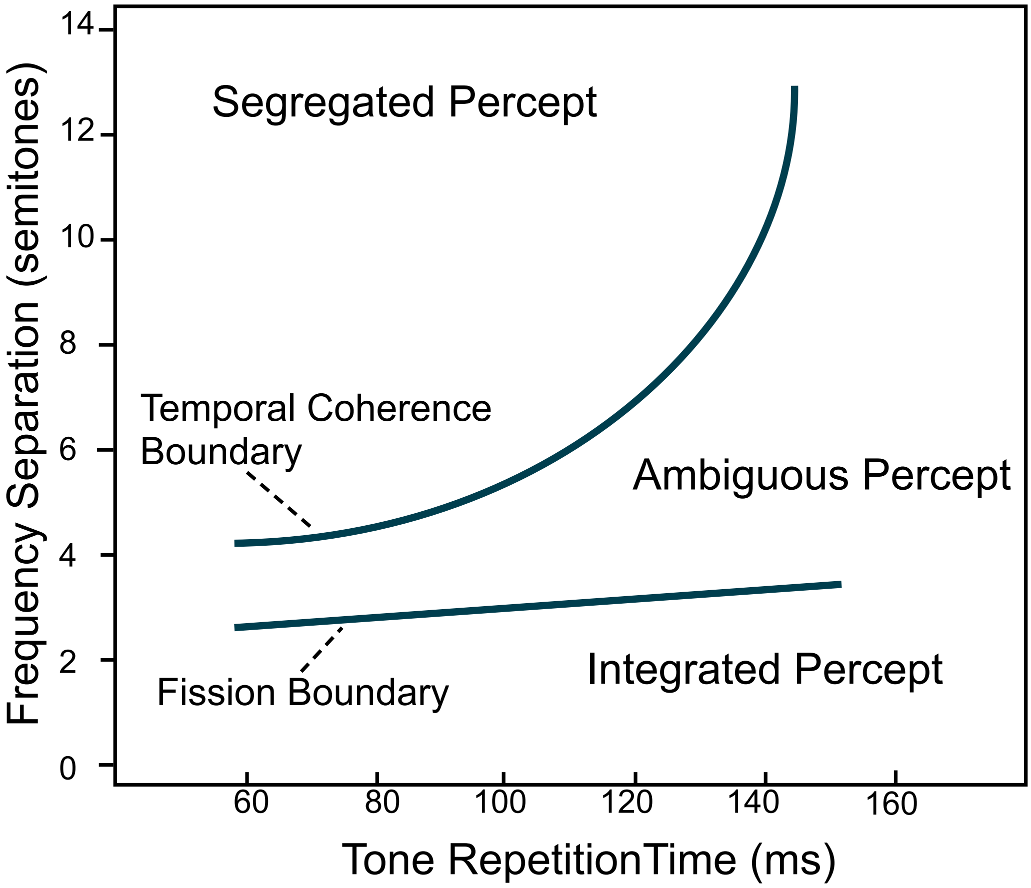


Figure 1.2: *The interaction between frequency separation and presentation rate on the perception of a repeating LHL- sequence. Frequency separation is measured in semitones, and presentation rate is measured in terms of tone repetition time (TRT = the onset-to-onset interval between successive tones). A short TRT corresponds to a fast rate and whereas a long TRT corresponds to a slower rate. Adapted from van Noorden (1975).*

In the frequency domain, it had initially been suggested that stream segregation primarily depends upon cochlear filtering where different sounds excite different regions of the basilar membrane in the cochlea of the inner ear (Hartmann & Johnson, 1991). The extent of streaming depends on the degree of overlap of the excitation patterns in the cochlea; a larger amount of overlap would lead to fusion or a single stream percept whereas a smaller amount of overlap would lead to fission or perceiving the sound sources as separate streams (Beauvois & Meddis, 1996;

McCabe & Denham, 1997). Hartmann & Johnson (1991) had previously proposed that stream segregation of two sounds was determined by parallel band-pass filtering, i.e., ‘channelling’ of incoming sounds by the auditory periphery (summarized as peripheral resolvability), especially based on their resolved fundamental frequency (f_0).

However, there is evidence that streaming can occur in the absence of spectral cues, especially in cases of segregation of complex tones. The spectrum of a complex tone contains a number of harmonics. Because the width of the auditory filters in the basilar membrane is, roughly, a constant proportion of their characteristic frequencies, the filters become broader as frequency increases. At high frequencies, components need to be further apart in frequency to be physically separated across the membrane. It is known that only about the first 10 harmonics in a complex tone are separated out or resolved by the cochlea, and excite distinct places on the basilar membrane (Plomp & Mimpen, 1968). A place on the membrane tuned to a low harmonic shows a pattern of vibration corresponding to the waveform of the pure-tone harmonic. In contrast, the higher harmonics are unresolved, and several harmonics interact at each place on the membrane. As the filters are broader towards the base of the cochlea, i.e. the higher frequency regions, higher harmonics of sounds (approximately 11th harmonic upwards) typically tend to be unresolved.

Psychophysical experiments carried out by Vliegen & Oxenham (1999) as well as by Grimault and colleagues (2000) showed that perceptual stream segregation for complex tones could occur even if the stimulus consists only of high frequency, unresolved harmonics. In both sets of studies, harmonic complex tones that were band-pass filtered to only include the higher harmonics were used. The studies found that listeners perceived stream segregation even in the high frequency band, which only consists of unresolved harmonics. These results indicated that sound sequences differing by f_0 could be segregated even when the f_0 was physically removed from the stimulus and the individual harmonic components of the complex tones were unresolved at the level of the basilar membrane. This set of studies along with other studies carried out in a similar vein (Cusack & Roberts, 2000; Roberts et al., 2002) highlighted that 'although resolvability of the harmonics is not absolutely necessary for streaming, it significantly contributes to the extent of segregation' (Grimault et al., 2000). Furthermore, these studies indicated that having distinct, spectrally separated stimuli was not a prerequisite for stream segregation and that stream segregation could be solely based on periodicity information available from resolved or unresolved components in the acoustic stimulus (Vliegen & Oxenham, 1999).

In their reviews of the literature regarding properties of stream segregation, Moore & Gockel (2002, 2012) concluded that auditory stream segregation is primarily based on any perceptual differences between sounds, which are not limited to cues obtained from peripheral frequency channels within the cochlea. They suggest that although spectral differences within sounds are a predominant cue for segregation, other physical aspects like differences in temporal envelope, temporal coherence and binaural inputs play a significant role in how the auditory scene is analysed.

1.2.2 Build-up of stream segregation

A distinct characteristic of stream segregation is that when listeners are exposed to a repeating sequence of low and high frequency pure tones that do not change in their acoustic parameters, the tendency to hear a segregated percept increases over time. This has been termed as 'build-up' of stream segregation. The first dedicated study of build-up of segregation over time was conducted by (Bregman, 1978). In this study, the length of a "tone package" (tone packages comprised either of 12, 24 or 48 repetitions of an alternating High-Low stimulus, where High and Low indicate pure tones with high and low frequencies) was varied and the perceived segregation was measured via subjective response. Each individual tone package was separated by four seconds of silence. In each trial, only one category of the tone package was repeated indefinitely. Listeners were instructed to increase the rate of the sequence, from an initial tone-repetition-time (TRT) of 600 ms, until the sequence was perceived as segregated. The rate at which segregation was heard was the measure of segregation - an increase in stimulus rate was known to promote stream segregation. Hence, the hypothesis was that if segregation was heard at a slower rate, it indicated the presence of another factor that promoted stream segregation. Bregman (1978) found that a segregated percept was heard at slower rates for longer tone package lengths. This means that even for slower rates where the temporal rhythm was within the acoustic parameters of an integrated percept, prolonged exposure to the sequence changed the listeners' initial integrated percept into that of a segregated percept. This finding provided evidence that as the number of tones in a package was increased, the tendency to perceive a segregated percept increased.

Anstis & Saida (1985) conducted a follow-up study to the Bregman (1978) study by directly measuring the listeners' integrated or segregated percept. In their paradigm, they presented long (30s and 60s), unchanging sequences of the 'High-Low' tone stimulus. In the first experiment, listeners continuously reported their perception of the tone sequences by holding down one of two keys (for integrated or segregated percepts) indicating either an integrated or segregated percept. Results suggested that for all tone sequences tested, the initial percept was that of integration. However, the tendency to report a segregated percept increased over the time course of the tone sequence.

This finding held true for all the tone sequences measured. However, the acoustic properties of the tone sequences affected the rate of build-up (of the segregated percept) as well as the extent of build-up. In follow-up experiments, the presentation rate or tone-repetition-time, TRT of the tonal sequence was varied. The results indicated that for the sequences that were more likely to be perceived as segregated (i.e., those with a rapid

TRT), initial build-up time was more rapid and the overall probability of reporting a segregated percept was higher (see Figure 1.3). Build-up of stream segregation was typically characterised by an initially rapid increase in segregation (i.e., within the first 5 - 10 s of a sequence), and a much more gradual increase (if any) over the remaining portion of the tone sequence. Depending on the rate of increase of streaming in the first few seconds of the sequence, build-up of segregation could be observed for a relatively short tone sequence, such as a sequence less than 5 seconds. A key thing to note is that Anstis & Saida (1985) analysed their subjective response data by averaging the responses across all trials. This is important to note as more recent studies that suggest a stochastic nature of build-up analyse the behavioural responses differently (as described further on in this section).

Recent studies have used a similar direct report of perception (by asking listeners to indicate integrated or segregated percepts), but have analysed the duration of the individual intervals of each integrated or segregated percept (Denham & Winkler, 2006; Pressnitzer & Hupé, 2006) separately. These studies calculated the average duration of each successive percept and reported that when listeners made subjective perceptual judgments on long tone sequences, a perceptual stage where the percept was exclusively that of segregation was never reached. Pressnitzer & Hupé (2006) presented a High-Low-High- tone sequence (High and Low being pure tones with a five semitone frequency separation) to listeners and found that in the average durations, there was no long-term trend for either of the two percepts (integrated or segregated). They showed that each percept, either integrated or segregated, lasted for approximately 9 seconds. The only exception to this percept duration was that the initial integrated percept was found to last longer than all subsequent percepts (approximately 17 s). They interpreted these findings as understanding the percept of an alternating tone sequence as being bistable.

However, a strong criticism of the study by Pressnitzer & Hupé (2006) was that only one particular type of tone sequence was investigated. This issue was addressed by (Denham & Winkler, 2006) who presented a four-minute long alternating tone sequence with various frequency separations. They found that the average duration of the first percept declined with an increase in frequency separation of the tones. These findings were consistent with the observations of Anstis & Saida (1985).

As seen from the various studies described above, the perceptual build-up of segregation has been an aspect of segregation that has not been understood fully. Furthermore, the physiological mechanisms responsible for the build-up of stream segregation are not well understood. Anstis & Saida (1985) attributed the build-up effect of streaming to frequency-shift detector units in the brain, which integrate successive tones into a single stream. With repetition of the stimulus, these detectors are suggested to habituate and the breakdown of their integrative function results in the formation of two separate streams. Other studies have attributed the mechanisms of build-up to the functioning of peripheral mechanisms (Beauvois & Meddis, 1996; McCabe & Denham, 1997). However, as noted earlier, peripheral channelling may not be a pre-requisite for stream segregation (Vliegen et al., 1999; Cusack & Roberts, 2000).

An explanation for the occurrence of build-up was put forth by Bregman (1990). He suggested that at the onset of a sound event, the default assumption is that all acoustic elements in the event arise from the same source. However, as the two subsets of sounds are found to have distinctly different properties, there is increasing evidence of the likelihood of two different sound sources based on the acoustic differences. The accumulation of this evidence towards the internal hypothesis of there being more than one sound source is a slow and relatively conservative process (in order to prevent excessive fluctuations due to insufficient evidence in the perceptual organization of a sound sequence). This functional explanation of what could be the reasoning behind the build-up of segregation can then explain the longer duration of the initial integrated percept (Denham & Winkler, 2006; Pressnitzer & Hupé, 2006).

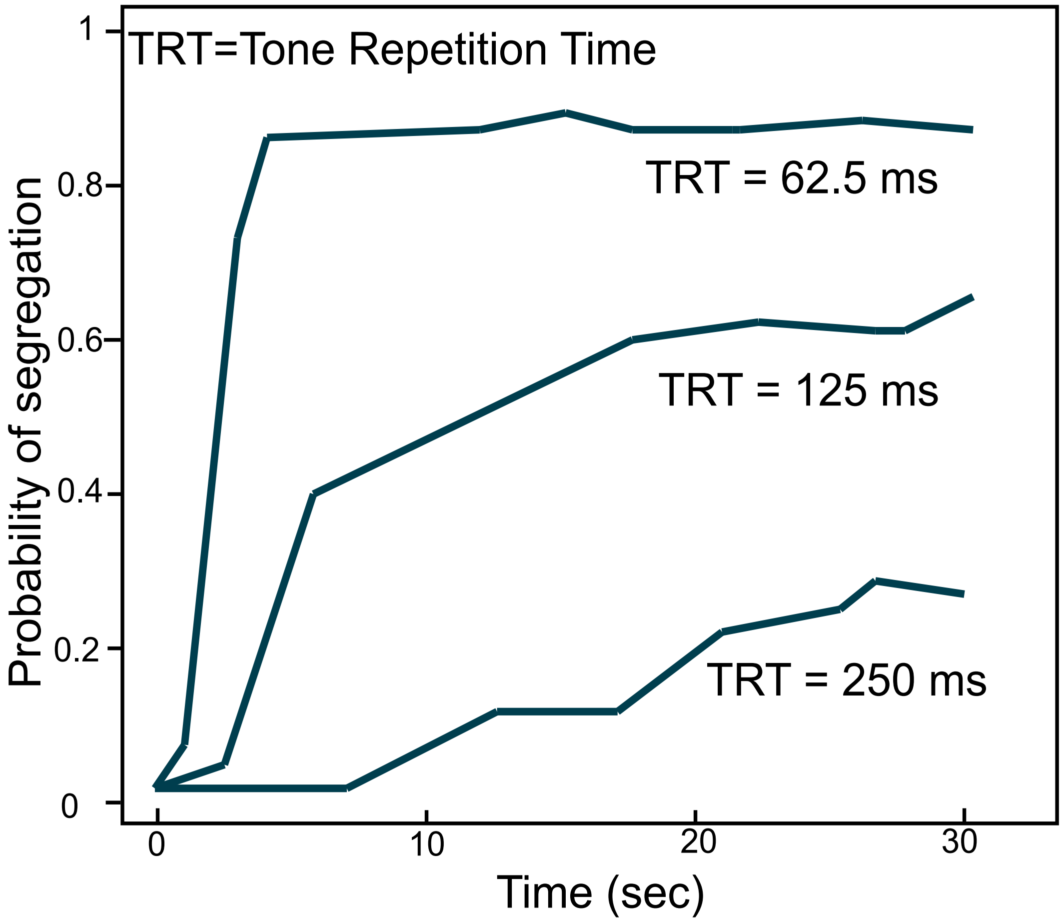


Figure 1.3: *Perception of listeners for various tone repetition times (TRT) for a sequence of 30 seconds showing different rates of build-up of segregation Adapted from Anstis and Saida (1985).*

1.2.3 Bistability and other multistable percepts

More generally, any stimuli with clear evidence of any salient difference between sounds that are sequentially presented could potentially lead to stream segregation (Moore & Gockel, 2002). Simply put, this means that when two or more sounds are presented to a listener, any acoustical difference (either spectral or temporal) is enough to contribute to a potential percept of segregation. However, for a certain range of intermediate frequency separations, in the ambiguous percept region (see Figure 1.2), the percept tends to switch or ‘flip’ between one and two streams (van Noorden, 1975; Moore & Gockel, 2012). This is often referred to as multistability or bistability.

Multistable perception may occur when a static physical stimulus is capable of inducing more than one subjective percept. This percept tends to be stable over short periods of time, but characteristically changes from time to time (for example, in the case of an auditory tone sequence with two alternating tones, the percept can flip between that of an integrated or segregated percept) (Moore & Gockel, 2012; Schwartz et al., 2012). Understanding multi-stability can help us understand how objects or sources in the environment are grouped according to specific characteristics to form a coherent representation of our surroundings (Schwartz et al., 2012).

Perceptual bi-stability has mainly been demonstrated in the visual domain for a large variety of perceptually ambiguous visual stimuli (for example, a Necker cube (Necker, 1832) or binocular rivalry (Helmholtz, 1925).

Leopold & Logothetis (1999) have defined the fundamental characteristics of perceptual bi-stability (across senses) as, 1) exclusivity (this suggests that perceptual representations are mutually exclusive and that only one perceptual representation of the ambiguous stimulus may be perceived at any given time), 2) inevitability (there will always be switches in the perception of the ambiguous stimulus) and 3) randomness (the successive durations of each of the percepts may be uncorrelated or stochastic).

Pressnitzer & Hupé (2006) have also suggested that these flips in auditory streaming paradigms do not occur in a fixed, regular manner. They suggest that when listeners perceive an ambiguous alternating High-Low frequency tone sequence, the percept follows the characteristics listed above suggested by Leopold & Logothetis (1999) for visual ambiguous stimuli. Pressnitzer & Hupé (2006) suggest that the percepts of either integrated or segregated are mutually exclusive (at any given point in time, participants either hear the tone sequence either as integrated or segregated), the duration of the percepts had no long term trend and that volitional control by the listeners (for example, directed attention to either one of the percepts) had an influence on the percepts but did not totally remove alternations or flips between the two percepts.

Another model proposed by Tong and colleagues (2006) using multistable stimuli in the visual domain suggested the idea of distributed competition. It has been suggested that when more than one plausible interpretation of a given stimuli is available, multistable perception or ‘stable instability’ (Zeki, 2004) is bound to occur. For example, in the case of an alternating tone sequence with tones that fall within the ambiguous parameter range, two plausible scenarios could exist: either the sounds emanate from one sound source or they emanate from two sound sources. However, in the case of a single pure tone repeating at a particular rate, the scenario of two sources producing the exact same tone is rather implausible. In this case there would be a very low chance of any ambiguity or instability.

1.2.4 Stimuli used to study auditory scene analysis

1.2.4.1 Tonal stimuli

One of the simplest stimuli used to study auditory streaming was originally suggested by van Noorden (1975) in the form of repeated presentations of auditory stimuli such as a tone triplet in the form of A-B-A_ (where A and B refer to tones of different frequencies) (Figure 1.1 illustrates a schematic of the classic A-B-A_ tone triplet paradigm where A and B are pure tones of different frequencies, which can either be perceived as an integrated or segregated percept). The extent of the difference in frequency between the two tones and the duration of inter-triplet and intra-triplet intervals can influence the streaming percept (inter-triplet intervals refer to the silent period between the repeating triplets of tones whereas intra-triplet interval refers to the silent periods between the component tones in each triplet). The triplets are repeated in sequences and can either be perceived as a galloping, 1-stream integrated percept or a 2-stream percept where the two different tones bind according to common factors (for e.g. frequency) and sound like two segregated independent streams. The factors that affect this stimulus have been discussed earlier in this chapter. The primary two factors that influence it are the frequency difference and rate of presentation (van Noorden, 1975). However, apart from spectral cues, non-spectral factors also influence stream segregation (Moore & Gockel, 2012). These cues include rate of fluctuation of temporal envelope (Grimault et al., 2002), timbre (Iverson, 1995), phase spectrum (Roberts et al., 2002), fundamental frequency (F_0) (Vliegen & Oxenham, 1999), lateralization cues such as interaural time differences (ITD) (Darwin & Hukin, 1999; Stainsby et al., 2011), onset and offset asynchrony (Darwin & Carlyon, 1995), harmonicity (Moore et al., 1986), and ear of entry (Darwin & Carlyon, 1995).

Another experimental paradigm that has been used to study stream segregation using techniques like electroencephalography (EEG) and magnetoencephalography (MEG) in audition is the 'oddball paradigm' (Näätänen et al., 1978). The paradigm is based on a two-tone sequence in which the probability of occurrence of the tones is varied; the standard tone is repeated more frequently and a few deviant tones occur in the stimulus sequence sparsely in a random fashion. The 'oddballs' in such a paradigm are the deviant tones. This paradigm has been used in both human and animal electrophysiological studies and can be used in passive listening conditions. Each of the tones in the sequence elicits an evoked response; the key being that the oddball deviants elicit a larger response than the standard tones. Traditionally, the evoked responses using EEG or MEG to the standard and the deviant tones were averaged separately and the difference between the two response waveforms (standard and deviant) was termed as the mismatch negativity (MMN). This paradigm has also been used in auditory scene analysis experiments where a differential response to the deviant tone (in the form of the MMN) can be used as a neural measure of the percept in scene analysis. For example, Sussman et al., (1999) presented alternating High and Low frequency tones to listeners in a passive listening condition. The paradigm was such that for the standard stimulus, if the alternating tone streams segregated into two distinct streams, repetitive tone patterns (repeating patterns of three rising tones or three falling tones) would emerge within each stream. The deviant trials did not include these repeating rising or falling tone patterns. Detection of these alternating pattern violations was expected to evoke an MMN as they deviated from the original alternating pattern. However, if the streams were not segregated into separate components, an MMN would not be elicited, as the unsegregated pattern would only be heard as a mix of different frequency tones. This shall be discussed in greater detail in section 1.2.6.

More recently, there has been a push towards using more naturalistic stimuli in scene analysis, which are richer in spectro-temporal characteristics as well as allow for greater flexibility in modulating different aspects of the auditory scene (Overath et al., 2008, 2010). This is essential as the canonical ABA_ tone triplet stimulus is not one often encountered in a natural environment.

1.2.4.2 Speech

Speech and conspecific vocalizations are real-world natural communication signals used by human and non-human animals. They are characterized by their rich spectro-temporal content. It has been suggested that speech is the perfect stimulus to use in human studies on the cocktail party problem (Cherry, 1953; Billig et al., 2013), as it is the actual stimulus that is found in this problematic scenario. However, since it is such an inherent property in human listeners, it is more likely to be contaminated by top-down processes like semantic properties, familiarity with the content and attention among other higher-level cues (McDermott, 2009). Speech signals are also not amenable to careful manipulation of their spectro-temporal properties unlike tonal stimuli described in section 1.2.4.1. Several psychophysical, imaging as well as electrophysiological experiments using speech in humans and con-specific vocalizations in animal models have been conducted. A few studies highlighting the use of speech stimuli in sound segregation are discussed in this section. One of the leading theories of speech perception suggests that the perception of speech depends on the entrainment of cortical activity to various temporal units in the speech signal, such as syllable and phoneme rates, which enables the listener to parse the speech signal into different units of speech at different frequencies (Giraud et al., 2007; Lakatos et al., 2008; Giraud & Poeppel, 2012). As a result, the neural activity related to speech encoding consists of multiple frequency bands at these individual frequencies. In a cocktail party scenario, selective entrainment to the temporal features of the attended speaker is important to track his or her speech over time. Several imaging studies have explored the question of attentional control in the context of a multi-talker environment using techniques with high temporal resolution such as EEG, MEG and intracranial EEG to track the precise temporal dynamics of speech (Lee et al., 2013; Bharadwaj et al., 2014).

Luo & Poeppel (2007) studied the MEG responses of listeners presented with spoken sentences. The resulting MEG responses were analysed for their frequency content as well as phase tracking dynamics. They found that the phase pattern of theta band (4-8 Hz) responses in the auditory cortex reliably discriminated spoken sentences. As this frequency range also correlates with the syllable rate of average spoken speech, they suggested that a temporal window corresponding to the theta range (approximately 200ms) segments the input speech signal and may be involved in processing syllables (mean duration of approximately 200ms). These results were further supported by findings that suggested that selective attention enhanced the discrimination of attended speech in the auditory cortex in a frequency range from 4-8 Hz (Kerlin et al., 2010).

Ding & Simon (2012) presented listeners with speech stimulus from two competing speakers (same or different genders) in an MEG experiment and asked them to attend to one of the two speakers while the relative intensity between the attended and the background speakers was manipulated. Their results suggested that the MEG responses showed the neural representation for the speech of both speakers, wherein each response was selectively phase locked to the rate (or rhythm) of the speech stream. This representation was accurate enough to enable the reconstruction of the original speech temporal envelope from the MEG signal. Furthermore, the robustness of this reconstruction was insensitive to the relative intensity of the attended speaker, i.e. even if the target speaker was presented at a lower intensity, the representation of the target speech was unaffected suggesting that a robust object-based representation of the attended speaker was formed.

A similar experiment by Mesgarani & Chang (2012) using multi-electrode surface recordings known as Electrocorticography (ECog) from the human auditory cortex showed that it was possible to reconstruct both the attended and the ignored speech signal from the time course of high-gamma power in the recorded neural activity. They also reported that the attended speech signal could be more reliably reconstructed than the ignored one. In the same vein, Zion Golumbic and colleagues (2013) also showed that both low frequency phase and high gamma power concurrently track the envelope of attended speech and suggest that tracking in these two bands may represent separate neuronal mechanisms for speech perception. Furthermore, attention was shown to modulate the perceptual representation of the speech signals in the auditory cortex by enhancing the representation of the attended speech stream, although the ignored speech stream remained represented as well. However, in higher-order cortical areas, the selective representation of the attended speaker was still robustly observed but no representation of the ignored speech was seen.

1.2.4.3 Complex non-speech stimuli

More recently, complex, synthetic auditory signals have been used to examine the cocktail party problem (Nelken, 2004). These auditory signals were designed to either replicate complex everyday auditory scenes while allowing precise flexible control over the acoustic properties of the stimulus in spectro-temporal space, which was not possible in recorded samples of real-life signals (Overath et al., 2010) or were designed to replicate the statistics of stationary sounds in the environment (McDermott & Simoncelli, 2011).

Overath et al. (2010) suggested that the analysis of objects comprised of two fundamental perceptual processes (Griffiths & Warren, 2004; Bizley & Cohen, 2013). The first mechanism involved the detection of boundaries between objects and was based on the identification of the variations in properties of different individual objects at the edges of these objects in the spectro-temporal domain (Kubovy & Van Valkenburg, 2001; Chait et al., 2008). The second mechanism involved having an unvarying representation and maintenance of the individual segregated objects (Griffiths & Warren, 2004). Studies have examined the cortical bases and representation of auditory edge detection (Chait et al., 2008); however, these studies do not assess the mechanisms relating to perceptual representation of the individual segregated auditory objects. Overath et al. (2010) developed a new stimulus based on spectro-temporal coherence that allowed flexibility in terms of creating objects and the acoustic boundaries or edges between them. The stimulus consisted of randomly distributed linear frequency-modulated ramps with different trajectories. The coherence between these ramps was manipulated to create different auditory objects and the transitions between ramps with different coherence represented boundaries between these objects.

A similar stochastic stimulus was developed by McDermott & Simoncelli (2011) that was based on capturing the statistics of real-world stationary sound textures such as a stream of water, the sound of fire or sound produced by a swarm of insects. In a recent experiment based on such textures, McDermott et al. (2013) developed a 'cocktail party' texture that was based on the superposition of multiple recordings of different speakers. Four different versions of the textures with varying density or number of speakers (1, 7, 29, or 115) were created. Listeners were presented with three excerpts of textures (of which two were identical) and were required to indicate which excerpt was different from the other two (as in an AXB paradigm). Two different durations of the textures were used: 50ms and 2500ms. Results revealed that the shorter stimuli were highly discriminable for all conditions but varied for the longer stimuli, producing an interaction between duration and the density of the textures. These results are in line with other experiments in the same study where discrimination of different exemplars of the same texture declined with the duration of the textures. This is contrary to discrimination performance for samples of different textures where performance increased with duration. Overall, the results suggest that summary statistics for mixtures such as speech may have a role in encoding time invariant properties of speech like voice quality or speaker identity and thus may aid segregation based on these features.

All of these stimuli have been used to study various aspects of sound segregation. It has been noted how speech or speech like stimulus has been used to greater extents recently to study segregation and its various properties. However, the issues with using speech as stimulus still remain (as discussed earlier) and several other properties and nuances of streaming still need to be understood. One important approach for further exploring the nuances of stream segregation is using ambiguous percepts (Pressnitzer et al., 2011) as discussed in section 1.2.3. Usually, studies have used ABA_ tone triplets in the ambiguous range to address this (Gutschalk et al., 2005). However, another category of stimuli that could be investigated as ambiguous stimuli are auditory illusions. Section 1.2.5

as well as section 1.5 shall highlight the potential of using auditory illusions to study stream segregation.

1.2.5 Illusions in stream segregation

It has been suggested in the early 70s that studying illusions and confusions, both visual as well as auditory, is essential as these supposed ‘failures of perception’ isolate and clarify the fundamental processes that conventionally lead to accuracy of perception and appropriate interpretation of ambiguous sounds (Warren & Warren, 1970). In effect, illusions help to establish new parameters within which one can study properties of established perceptual processes. With respect to auditory stream segregation, two distinct illusions suggested by Diana Deutsch; the Octave Illusion and the Scale illusion (Deutsch, 1974; 1975) tap into different schema-based integration in streaming (Bregman, 1990 - see section 1.1).

Both the octave and scale illusion have multiple tone sequences that are temporally coherent across ears. However, their illusory percepts do not follow the exact rules of streaming as studies using ABA_ paradigms and other conventional streaming paradigms. Details of the illusion will be described in section 1.5. However, the key need for exploring how illusions work is to help understand in what ways the auditory system segregates sounds when the parameters of a stimulus are indiscriminately linked together. At this stage, there must be a set of mechanisms that enables the listener to form linkages using other principles between some elements that go beyond the traditional streaming links that have been studied until now.

The octave illusion has not been intensively studied from a stream segregation point of view. However, one could argue that it could lend itself as a tool to study the interaction of synchronous and sequential sound segregation, as the structure of the illusory stimulus consists of an ABAB_ sequential streaming pattern in each individual ear as well as having the tone streams in each ear presented synchronously at any point in time (see Figure 1.4). Lamminmäki & Hari (2000) have briefly indicated that stream segregation may play a role in the way in which the component tones are parsed. However, in a more generic context, the illusory stimulus provides an interesting problem to study auditory stream segregation for alternating as well as synchronous tones while using only two tonal components.

1.2.6 Response measures for streaming

1.2.6.1 Psychophysics

Psychophysics has been the primary experimental technique to study stream segregation. Most of the studies mentioned in the previous sections have used some form of psychoacoustic measure. Seminal work to establish spectro-temporal parameters of stream segregation as well as distinct characteristics such as build-up of segregation has primarily involved psychophysics (Anstis & Saida, 1985; Bregman, 1978, 1990; Bregman et al., 2000; Bregman & Campbell, 1971; Moore & Gockel, 2002, 2012; van Noorden, 1975). The effect of top-down factors such as attention has also been studied using various psychoacoustic paradigms (Carlyon et al., 2001; Cusack et al., 2004; Macken et al., 2003).

There is no one correct or standard method of measuring behavioural outcomes of streaming. However, there has been an interesting gradual shift in the way outcome measures of streaming have been considered over time. The early studies of streaming used the method of adjustment where the listeners were asked to adjust the properties of the stimuli (Anstis & Saida, 1985; Bregman, 1978) until there was a change in percept. Another approach used is the 'method of limits' (Anstis & Saida,

1985) where the examiner slowly manipulates the properties of the stimulus and the listener indicates when there was a perceptual shift. For example, in a paradigm with a repeating ABA_ tone sequence, the examiner may slowly increase the frequency separation between the two tones. The listener would have to indicate when the percept shifts from that of one-stream to a two-stream percept. A modified version of this method includes calculating the proportion of integrated and segregated time where the participant holds down a button for a particular percept till there is a change in percept (Gutschalk et al., 2005; Micheyl & Oxenham, 2010; Pressnitzer & Hupé, 2006). This would require the listener to report their percept throughout the stimulus rather than just reporting the change in percept. The total duration of the different percepts are then averaged to get a proportion of time that the two different percepts were heard.

Over the last decade, there has been an increased focus on objective measures of streaming. Several difficulties have been noted in the interpretation of subjective responses as the ones resulting from the methods described above. The first difficulty that arises from the lack of objective criteria is that it becomes very difficult to determine whether the listeners follow the instructions given to them, or in cases when no specific instructions are given, there is very little control over what the listeners are listening out for (Micheyl & Oxenham, 2010). The second difficulty arises from the fact that subjective responses such as one-stream or two stream responses (or integrated/segregated responses) are predisposed to individual biases, which could either be sensory (for example, listeners may have an a priori inclination to try to integrate or segregate) and/or may also be decisional (for example, different listeners may have different criterion for responding “two streams”, etc.). These limitations need to be taken into account when subjective and objective psychophysical measures of auditory streaming are compared or correlated.

As opposed to the previously mentioned subjective methods where the participants had to report their perceptual response with regards to streaming (for example, reporting whether one or two streams were heard), objective measures of streaming are obtained using psychophysical tasks where perceiving the stimulus as an integrated, one-stream percept is advantageous to the performance on the task, or an opposite condition where segregation is key to better performance on the objective task (Micheyl & Oxenham, 2010). For example, a temporally shifted tone can be a target deviant in an ABA_ tone paradigm. A temporal shift is easier to detect in an integrated percept than a segregated percept (Micheyl & Oxenham, 2010). The performance on such a deviant detection task is used as an objective measure of the streaming percept. This method has been used in several studies of segregation to provide for a more concrete measure of the percept of segregation (Thompson et al., 2011; Szalárdy et al., 2013).

1.2.6.2 Electrophysiological studies in non-human animals:

Stream segregation in animals has been studied for both single-unit and multi-neuronal setups in several organisms including macaques, ferrets, European starlings and zebra finches among others (Fishman & Steinschneider, 2010). Seminal work in streaming was done using macaques by Fishman and colleagues (2001) in primary auditory cortex (A1) who studied the relationship between sequential streaming and neural responses to temporal sound sequences by recording multi-unit neural activity and current source density in response to alternating-frequency tones. They found that at slow presentation rates of sequences with alternating A and B tones, cortical sites show a prominent neural firing to both A and B tones, even when the recording site has a best frequency (BF) response to A tones. However, at faster rates, the response to the B tones is suppressed and there is only a marked response to the BF ‘A’ tones. Furthermore, the degree of suppression of the B tones (non BF tones) was directly proportional to the amount of frequency separation. These results have been explained in terms of ‘physiological forward masking’ where the neural response to a stimulus is reduced due to the influence of a preceding stimulus. The proposed neural model for streaming based on these findings suggested that in A1, the responses to each of the tones (A and B) are represented in areas along the tonotopic axis (segregated according to frequency where separate regions along the axis respond to different frequencies). As these two get separated, they contribute further to the percept of two distinct sound sources. They suggest that the key factors that play a role in this cortical model are adaptation (related to rate of presentation), frequency selectivity and forward masking. The drawbacks of this model are its inability to explain streaming of sounds with overlapping spectral properties as well as the fact that it does not take any non-primary auditory cortical areas into account (areas apart from A1).

Another important study of animal streaming carried out by Micheyl and colleagues (2005) on awake monkeys used single unit recordings to long ABA_ sequences at different frequency separations and compared the neural firing rate responses to psychophysical subjective responses obtained from human listeners who were instructed to indicate the number of streams heard for similar presentations of long ABA_ tone sequences. Based on their neural findings and by confirming that they mirror human perceptual results, they proposed a model based around the idea of neurons acting like binary classifiers. They suggested that neurons in the primary auditory cortex read out the spike counts received for each of the tones in the streaming triplet. This classification is predicted to be based on measures of spike counts evoked by the A and B tones in a streaming triplet. If the spike count for only one of the tones exceeds a specific threshold, the tone is deemed segregated however if both the tones elicit spike counts that exceed a particular threshold, the sounds are deemed to originate from a single source.

Streaming studies in songbirds like the European starling have also been carried out for natural sound signals like conspecific birdsongs as well as synthetic tonal signals (MacDougall-Shackleton et al., 1998; Bee & Klump, 2004, 2005). The findings in avian species replicates the findings in macaques and backs up the role of forward masking, adaptation and frequency selectivity in the cortex for processing streaming stimuli.

1.2.6.3 Neuroimaging studies in humans

The past few decades have seen a large number of neuroimaging studies in humans using various stimulus paradigms and various imaging techniques like EEG, MEG and fMRI to study auditory streaming.

Neuroimaging studies in humans focusing on streaming have more commonly used the mismatch negativity (MMN) paradigm to study various aspects of stream segregation (Sussman, 2007; Sussman et al., 2005; Winkler et al., 2003). The MMN is commonly thought to be a result of a pre-attentive process as it can be evoked in sleep, anaesthesia or minimal states of consciousness (e.g. Boly et al., 2011). However, under certain conditions, it can also be modulated by attention (Alain & Woods, 1997; Sussman et al., 2003). Sussman et al. (1999) used the MMN, suggested to be a neural index of pre-attentive acoustic processing, to study the role of attention in stream segregation. They presented sequences of six different high and low frequency tones in fast and slow rates of presentation conditions. Overall, they found that an MMN is elicited in the fast paced condition, where streaming is said to have occurred, for both high and low tones indicating that stream segregation may be initiated at a pre-attentive state. Although attention is not a prerequisite for an MMN to be evoked, it is known that the MMN is modulated by attention (Alain & Woods, 1997) and higher-level influences on stream segregation have been established. However, from the study by Sussman et al. (1999) as well as several other studies using MMN in streaming paradigms, it can be established that there is a critical pre-attentive role in the formation of auditory streams.

As noted, the MMN has been used extensively to study stream segregation. However, the MMN, as the name suggests, is evoked by a mismatch paradigm and 'requires events that that could be only perceived as deviants if streaming has occurred' (Snyder et al., 2006). For example, Winkler et al. (2003) presented synchronous sequences of sounds; for one sequence of monotonous tones, the intensity of the tones was kept constant apart from occasional intensity deviants (target deviants) and another sequence where the intensity of the tones were constantly varied. In the case when the sequences were perceived as an integrated percept, the target occasional deviants were not considered deviants as they were grouped with the constantly changing intensities of the tones of the other stream. However, when perceived as a segregated percept, the occasional deviants elicited an MMN.

Even though the MMN enables us to see if streaming has occurred, it does not give us too much information about the underlying neural mechanisms of segregation as there is no possibility of tracking the on going processing of the stimulus sequence (Snyder et al., 2006; Thompson et al., 2011). As stream segregation is known to vary according to stimulus length and exposure duration, having a measure that can be studied across the duration of the stimulus is crucial rather than having an aggregate response measure like the MMN. An alternative paradigm using evoked responses to study the relative amplitudes of the P1-N1-P2 peaks evoked by each tone in the stimulus was suggested by Gutschalk et al. (2005) and Snyder et al. (2006). Using the P1-N1-P2 evoked complex (using EEG or MEG) of tones to study streaming and effects of attention in bistable stimuli has been increasingly used as it can track the on-going processing of the stimulus (Snyder et al., 2006; Szalárdy et al., 2013). All of these experiments are based on the relationship noted between streaming and neural adaptation (Fishman et al., 2001). This hypothesised relationship is based on the auditory evoked N1m (MEG equivalent of the N100 in EEG), which is an evoked negative potential that usually occurs in the auditory cortex 80–150 ms after stimulus onset. The N1m is evoked for every distinct auditory stimulus. For a sequence of tones that have the same frequency (for example, a repeated tone sequence of tones at 1000

Hz), the N1m wave evoked by each successive tone adapts to a steady-state value that depends on the inter stimulus interval between the tones (Ritter et al., 1968). Furthermore, for a sequence of tones alternating in frequency and where the frequency separation between the tones is small, the N1m amplitude and its dependence on inter-stimulus interval remains the same as that for same frequency tone sequences, as long as the frequency difference between tones is small. However, for larger frequency separations, the amplitude of the N1m wave (and the corresponding degree of adaptation) becomes consistent with the longer inter-stimulus interval between successive tones of the same frequency rather than the shorter inter-stimulus interval between temporally adjacent tones of different frequencies. This demonstrates that adaptation occurs based on the frequency of the tones (Picton et al., 1978; Näätänen et al., 1988). This relationship of streaming and neural adaptation can be used to gauge the perceptual state of a listener in a stream segregation paradigm. This method allows us to track the evoked response to the on-going stimulus and may help us understand the development of streaming as the stimulus sequence progresses. Gutschalk et al. (2005) measured auditory evoked neuromagnetic fields using MEG in response to an ABA_ tone triplet paradigm in two separate experiments where the stimulus parameters were chosen to promote either a clear integrated/segregated percept or a bistable percept respectively using frequency separations that were either in the bistable range or easily separable range. The first experiment showed that changes in frequency separation and rate affected the magnitude of the auditory evoked responses that correlated with the degree of perceived stream segregation, i.e., the magnitude of P1m and N1m evoked by the B tones in the repeating triplet increased with larger frequency separations. This trend was also observed in the behavioural data (participants were asked to report their percept of the stimulus: either one or two streams) where high correlations were found between the magnitudes of the P1m and N1m evoked responses and the reported ease of streaming. The second experiment, where an ambiguous percept was induced showed results similar to experiment 1: the magnitude of P1m and N1m covaried with the perceptual state and the amplitude of the N1m was larger for two vs. one-stream percepts. Similarly, Snyder and colleagues (2006) observed that auditory evoked potentials, specifically the P2 and N1 in response to an ABA_ streaming stimulus sequence increased in amplitude with increasing frequency separation and correlated with behavioural measures of streaming. Furthermore, a slowly rising positivity was also found through the course of the sequence whose time course varied similarly to the build-up of streaming.

1.2.7 Computational models of stream segregation

Theoretical models of streaming vary in the properties of stream segregation that they take into account. Early models of streaming focused heavily on peripheral spectral analysis. An initial model suggested by Beauvois & Meddis (1991) suggested a ‘leaky’ integrative system that took into account peripheral frequency analysis, an attentional component and internal noise associated with the system. More simply, the model incorporated the following principles: 1) peripheral spectral analysis with band-pass frequency channels, 2) inherent system ‘noise’, 3) a principle of ‘leaky integration’ that would describe slow build-up and decay within the channels and 4) an attentional component that enables the listener to selectively respond to the maximally active channel. A later version of their model (Beauvois & Meddis, 1996) also included an adaptive component based on the responses of the auditory nerve. A modified version of the model suggested by McCabe & Denham (1997) included an inhibitory feedback system to explain the graded inhibition of either the foreground or the background signal. These models, however, fail to explain some nuances associated with streaming, due to them being fundamentally based on peripheral channelling. (Denham & Winkler, 2006) suggested an alternative view of modelling stream segregation by thinking of the auditory system as a predictive system involving several hierarchical levels. Their model suggests that the auditory system forms predictive models based on initial bottom-up physical segregation cues and also creates alternative models using top-down mechanisms. They suggest that there is on-going competition between these mutually exclusive perceptual models and depending on factors like focused attention or previous cues, a listener can perceive the auditory percept associated with the winning model at any given point of time. This type of model can explain factors like bistable percepts in segregation as well as take into account the top-down influences on streaming. A refined version of this model was suggested by Mill et al. (2013), which was again, based on having competing prediction of different aspects of the auditory scene, and has been able to explain several nuances of stream segregation like switching between percepts, build-up of segregation and the expected rate of switching between percepts.

An alternative view on streaming works along the lines of temporal coherence that was suggested by Shamma et al. (2011), which hypothesises that 'temporal coherence' between tokens in an auditory scene is an extremely strong cue for perceptual organization. Sources that are temporally coherent are strongly perceived as a single source and a lack of temporal coherence between two sound sources helps segregate them into separate streams. This model questions the previous models based mainly on spectral channelling by demonstrating that sounds that are distinctly spectrally separated are perceived as a bound single source if made synchronous/coherent (Elhilali et al., 2009). The temporal coherence model suggests that for sounds to be perceived as segregated, they need to differ in feature space (for example, spectral feature differences) as well as be temporally incoherent. It does not negate the need for frequency tuning and selectivity; in fact the first stage of the model consists of a filter bank that puts the incoming signal through a bank of band-pass filters. However, the next stage of the model analyses the temporal coherence between each of the output frequency channels and performs a dynamic correlation between pairs of channels to determine a coherence matrix and based on the degree of coherence, the perception of the sound signal can vary between integrated and segregated perceptual streams.

1.3 Role of attention in build-up towards a segregated percept

There is an on-going debate about the role of attention in the build-up of stream segregation. Bregman (1990) has suggested that stream segregation can either be stimulus driven, which he termed as ‘primitive processes’ or occur as a function of volitional attention control, which he termed as ‘schema–driven processes’. There have been two distinct groups of opinion as far as the role of attention in streaming is concerned. Some studies have argued that attention is crucial for build-up of streaming even when the process is mainly stimulus driven (primitive processes) (Alain & Woods, 1997; Brochard et al., 1999; Carlyon et al., 2001) whereas some studies have suggested that primitive stream formation can occur without focused attention (Bregman, 1990; Macken et al., 2003; Sussman & Winkler, 2001; Sussman et al., 2003; 2007).

Carlyon et al. (2001) investigated the role of attention in the build-up phase of stream segregation. They presented a 21-sec stimulus sequence containing repeated ABA_ tone triplets to one ear of the participants. The stimulus could either be perceived as a 1-stream (galloping) or two-stream (segregated) percept. They used four conditions with different frequency separations for the test stimulus. In the contralateral ear, series of noise bursts were presented and the listeners were instructed to judge each noise burst as either continuously increasing or continuously decreasing in amplitude. After 10 seconds of attending to the noise bursts, the listeners were instructed to switch their attention to the tone triplet stimulus sequence in the opposite ear. Immediately following the switch, the listeners had to make a subjective response via button press to indicate whether they perceived one or two streams. The noise-burst task was equivalent to a distractor task as the participants were to attend to the noise burst streams for the first 10 seconds of the stimulus while ignoring the tone sequence in the contralateral ear. The reasoning behind this experiment was that if build-up occurred without active attention, then the listeners should judge the tone triplet sequence as segregated as they have had 10 seconds to build up the segregated percept. However, if attention were needed for build-up, they would tend to respond as if they perceived a 1-stream percept due to lack of attention. They found that their results indicated that the listeners perceived the stimulus as a single stream after the switch and concluded that attention is required for the formation of auditory streams.

A major concern with this experiment has been that it does not take into account the streaming resetting that occurs when attention is shifted from one ear to the other (Carlyon, 2004; Cusack et al., 2004; Moore & Gockel, 2002). Active resetting of stream segregation due to switching attention between ears has been shown in various studies (Cusack et al., 2004; Roberts et al., 2008).

Macken et al. (2003) used a different paradigm incorporating the concept of the 'irrelevant sound effect (ISE)' in which sounds presented to the listener that were irrelevant to the listeners' task (in this case, a repeated series of alternating tones with different frequencies) disrupted performance on a visual recall task (Jones et al., 1999). Studies on these types of paradigms have shown that alternating tones perceived as one-stream are more disruptive towards performance on the visual task than two separate streams (Jones et al., 1999). Macken et al. (2003) presented alternating tones at 3 different rates as the irrelevant tones (or background distractor tones) while the participants were performing a visual recall task. They found that increasing the rate of stimulation disrupted performance until a point where the tones split into 2 separate streams due to build-up. This finding is in line with the findings of Jones et al., (1999), which suggest that segregated streams are much less disruptive to a visual recall task. Hence, they concluded that build-up was present even without focused attention on the auditory stimuli.

This paradigm has again been criticized on the basis of the task being a 'low load task' (Lavie, 2005) due to which the listeners have enough time to 'covertly attend' (Sussman et al., 2007) to the auditory stimuli. Also, since they base their outcome measures on the latter part of the stimulus, they cannot rule out the possibility that attention is only needed for the initial formation of streams (Sussman et al., 2007).

Sussman et al. (2007) conducted a series of experiments to test whether the build-up phase requires attention using the mismatch negativity (MMN) component. An MMN component is only generated when the sound sequences segregate into two streams (Sussman et al., 1999) and this was used as a neural indicator of stream segregation in these experiments. Their overall results indicate that attention is not always needed for the build-up of segregation, however, the spectro-temporal characteristics of the competing sound also govern whether segregation occurs for unattended sounds or not. They also further comment on the role of attention in segregation stating that even though they found that attention is not always needed for segregation, one cannot rule out that attention plays a role when the spectro-temporal characteristics of the stimuli fall in the ambiguous range of stimuli (Snyder et al., 2006; Sussman et al., 2007).

Thompson et al. (2011) carried out an experiment to study the effect of attention on build-up of segregation using a temporal deviant detection task, where one of the B tones in an ABA_ tone triplet paradigm was temporally shifted. They mainly compared two conditions: one where the listeners were instructed to detect a deviant late into the sequence and they were instructed to attend to the entire sequence and the other where they had to detect a late deviant after switching their attention midway towards the test sequence. The results indicated that participants who attended throughout to the stimuli could not avoid the build-up of segregation and hence performed worse in the temporal detection task as compared to conditions where the participants switched their attention.

Snyder et al. (2006) used an evoked potentials (P1-N1-P2) paradigm to study the effect of attention on stream segregation using the canonical ABA_ tone triplet paradigm. Their results suggest that the P200 evoked potentials increase in amplitude for trials where the listeners were instructed to focus their attention on the auditory stimulus as opposed to trials where they were instructed to ignore the stimuli (i.e. passive listening). However, they found that the EEG activity picked up automatic stream segregation of the sounds in the passive listening condition as well. However, the build-up process and amplitude of the evoked potentials was highly modulated by attention. Overall, they suggest that their experiment provided evidence for two mechanisms, one being the automatic segregation of sounds and the other being an attention dependent build-up of segregation that is heavily modulated by attention (Snyder et al., 2006).

1.4 Sequential versus concurrent sound segregation

Streaming has been studied in great detail for sequentially occurring sounds; however in a real world auditory scene, sounds rarely occur in isolation at any given instance in time. Thus, studying concurrent sound segregation becomes an important aspect of streaming that needs to be understood. However, studying concurrent sound segregation is not as obvious as one would imagine as the degree of concurrence or overlap plays a crucial role in the resultant percept. Detecting a mistuned harmonic in an otherwise periodic harmonic complex has been suggested as an easy example of understanding sound segregation (Moore et al., 1986; Alain et al., 2001). Bregman & Rudnicky (1975) initially studied the amount of temporal overlap needed for two tones to be perceived as either integrated and fused and found that the tones were perceived as a fused complex tone (i.e. not pure) when the overlap was greater than about 88 per cent. Also, with an increase in frequency separation, the segregation was clearer. Apart from overlapping pure tones and mistuned harmonics, another paradigm that has been highly informative in studying concurrent sound segregation is the double vowel task where two synthetic vowels are presented at the same time and listeners are supposed to report which two vowels were heard. Results from these studies have shown that the rate of correct identification increases with an increase in the frequency separation of the fundamental frequencies of the two vowels (Assmann & Summerfield, 1990, 1994; Chalikia & Bregman, 1989, 1993). In recent years, there have been a number of human electrophysiological studies using EEG and MEG that have identified responses to concurrent sound segregation in the primary auditory cortex (Bidet-Caulet et al., 2007) as well as non-primary auditory cortex (Alain, 2007; Lipp et al., 2010). Taking recent human behavioural studies (Micheyl et al., 2013) as well as previous streaming studies that have studied concurrent sounds in streaming (Shamma and Micheyl, 2010; Elhilali et al., 2009; Bregman, 1990), one can establish that synchronicity plays a distinct role in how acoustic sounds and streams are perceived.

1.5 Auditory Illusions

As discussed in section 1.2.5, illusions can be used as a method of understanding ambiguous stimuli and ways in which they can be segregated or perceived in the auditory and visual senses. In this section, auditory illusions suggested by Diana Deutsch (1974; 1975), namely the octave illusion and the scale illusion have been described.

The Octave illusion was initially elicited with a stimulus configuration consisting of two tones, spaced an octave apart, presented in an alternating Low-High tone pattern in different phases in both ears such that if the sequence in the Left ear started with a Low tone, the sequence in the right would start with a High tone. The resulting precept was an unexpected illusory percept where listeners perceived all the Low tones in one ear at half the presentation rate alternating with the High tones in the other ear at half the rate (Deutsch, 1974) (see Figure 1.4).

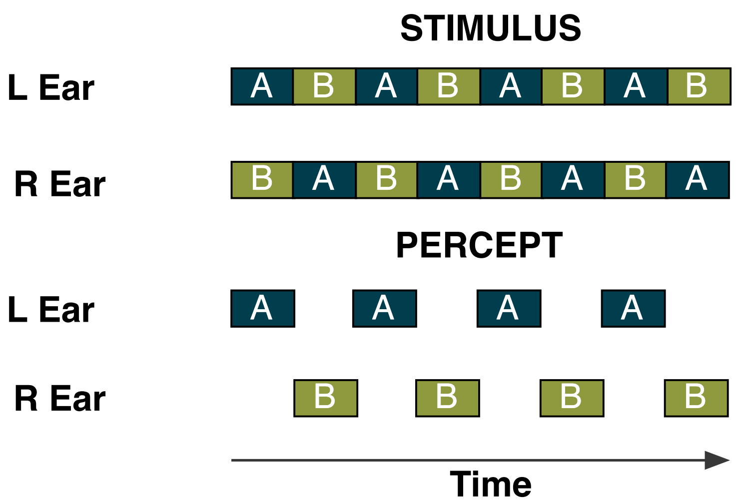


Figure 1.4: *Schematic representation of the stimulus and resulting illusory percept in Deutsch's Octave Illusion. Tones A and B indicate pure tones of frequencies 400 and 800 Hz.*

The Scale Illusion (Deutsch, 1975) was developed from the Octave illusion and was principally used to understand how simultaneous streams of musical information are separated into channels. It consists of a major scale with successive tones alternating from ear to ear. The scale is played simultaneously in both ascending and descending form; however when a tone from the ascending scale is in the right ear a tone from the descending scale is in the left ear, and vice versa. The resulting percept, however, is of a rising pattern in one ear and a simultaneous falling pattern in the other ear (Deutsch, 1975).

Both these illusions tap into various processes which could possibly involve stream segregation, as both the percepts require the listener to parse out individual tones from the complex stimulus and group it according to some pre-determined perceptual rules. The stimuli used in the experiments from Chapter 3 onwards are based on a variant of the octave illusion. Hence, the following review of literature is mainly focused on the octave illusion but many parallels can be extended to the scale illusion.

1.5.1 Initial discovery and theories of the octave illusion

The illusory percept associated with the octave illusion (as described above and in Figure 1.4) is a puzzling auditory phenomenon. Using established theories of stream segregation, it is not easy to explain both the perceived pitch and localisation. Furthermore, the properties of this percept gets more enigmatic considering that the subjective percept for listeners remains the same even when the headphones are reversed; i.e. the percept in Figure 1.4 would remain constant even if the stimulus sequence presented to the right and left ears remains the same.

The illusion presents a contradiction for theories of sound localisation and pitch perception. The alternating percept of pitch can be explained if one assumes that the listener attends to one ear and ignores the other. However, in that case, all the sounds would be heard in one ear. Alternatively, if the listeners' percept is driven by them attending to each ear in turn, the pitch should not change. However, the illusory component of this percept is that the low tone is localised in the ear that is physically receiving a high tone at that moment in time (Deutsch, 1974). Hence, this illusory percept cannot be explained by either of these theories in isolation. Deutsch (1975) proposed a dual-mechanism model to explain this illusion that consists of two separate brain mechanisms, one for pitch determination and the other for sound localisation that coexist and converge to elicit the illusory percept. The basis of this theory stemmed from work that suggests that there is indeed an anatomical separation in the auditory system between the sub-serving mechanisms of pitch and location determination (Poljak, 1927; Schneider, 1969). The model suggested that in order to determine the perceived pitch, listeners' attend to the pitch information in one ear and suppress the corresponding pitch information in the other ear. The second half of the model determines the perceived location of the tone based on the ear that received the higher frequency signal at that particular point in time. The model suggests that the final illusory percept is a combination of the output of the two sub-serving mechanisms.

Chambers et al. (2002; 2004) challenged the dual-mechanism model of Deutsch (1974) explaining the octave illusion. From their experiments based on subjective responses, they concluded that the perceived pitch difference between the alternating tones corresponded more to a semitone than to an octave. They also further suggested that participants do not consistently lateralize the tones towards the ear receiving the higher frequency, which is what is described in Deutsch's dual mechanism model. Based on their results, Chambers et al. (2002) proposed that the octave illusion percept is a result of diplacusis. Their explanation based on their findings suggested that listeners perceptually fuse the concurrently presented dichotic tones at any given point in time and the perceived pitch roughly corresponds to the fundamental frequency. The alternating percept is basically heard as two similar sounding tones (probably a semitone apart perceptually) where they attribute the difference in pitch to diplacusis. They finally suggest that because the perceived tones are the result of a fused percept, they can be lateralized in either ear and are not limited to being localised in the ear receiving the higher frequency.

1.5.2 Properties of the octave illusion

The stimulus used initially by Deutsch (1974) to elicit the octave illusion has been studied in different contexts and the robustness of the percept has been investigated across a variety of parameters. It has already been demonstrated that the percept of this illusion is robust to changes in tone duration (Zwicker, 1984), intensity (Deutsch, 1978), frequency separation (Brancucci et al., 2009), and timbre (McClurkin and Hall, 1981), and can also be elicited by aperiodic stimuli like band-pass noise (Brännström and Nilsson, 2011). It was noted by Deutsch and Roll (1976), and later confirmed by Brancucci et al. (2009), that the illusion is not dependent on the tones being in exact octave relationship. In fact, Brancucci et al. (2009) reported that the illusory percept was dominant for all musical intervals tested that were larger than a perfect fourth (roughly a ratio of 4:3 or a frequency difference of 33%). Deutsch and Roll (1976) suggested that listeners generally reported the tone frequencies that were presented to their “dominant” ear (usually the right), through suppression of the non-dominant ear. Such suppression was postulated not to occur for sound localisation, but instead the higher tone tended to be localised to the right, and the lower tone localised to the left, regardless of the ear of presentation.

1.5.3 Neuroimaging studies of the octave illusion

Few neuroimaging studies have been carried out to understand the octave illusion.

Lamminmäki and Hari (2000) carried out a study using MEG to investigate the octave illusion and divided their responses of interest into the transient responses and a sustained response. They proposed that the differential inter-hemispheric balance of the transient responses generated by the sound onsets determined the perceived location of the sound. They also suggested that the sustained response in the evoked EEG reflected the pitch perceived and that the low and high pitch sounds were monaurally separated using stream segregation in each ear. Lamminmäki et al. (2012) further used frequency tagging (modulating tones with specific frequencies in order to label the evoked activity associated with the particular component tone) in an MEG experiment and reported that their result suggesting that their results support the dual-mechanism model proposed by Deutsch (1975) (this study is described in detail in section 5.1).

However, in all the neuroimaging studies that aimed to study the octave illusion stimulus (see 1.4), either listeners' spontaneous percepts were tested beforehand and neuroimaging recordings were passively done without instructions on what to attend to within the sound sequences (Brancucci et al., 2012; Lamminmäki et al., 2012; Lamminmäki & Hari, 2000) or the response measures for a task based study were mainly focused on the listeners' subjective responses regarding their percept (Brancucci et al., 2014). The caveats regarding using purely subjective measures have been highlighted in section 1.2.6.1.

1.6 Outline of the thesis

The studies reviewed so far indicate that auditory scene analysis has been a topic of intense investigation over the last several decades. The effect of attention on scene analysis as well as the processing of multi-stable stimulus has been studied in several scenarios but there is still a need to test the principles of segregation with new stimuli that may challenge some models of streaming.

Chapter 2 describes an initial set of experiments that investigate the effect of attention on build-up in a classic, bistable streaming paradigm (ABA_ tone triplets introduced by van Noorden, (1975)).

Chapter 3 introduces a stimulus similar to the octave illusion (Deutsch, 1974) that has previously not been studied from an auditory stream segregation perspective. This chapter describes a set of experiments that were carried out to explore the effect of attention on the new stimulus as well as investigate whether stream segregation played a role in the perception of this illusory stimulus.

Chapters 4 and 5 delve deeper into understanding the mechanisms of the octave illusion.

Finally, Chapter 6 introduces a new model put forth to try and explain the mechanisms of the octave illusion based on the findings of the studies in chapters 3, 4 and 5.

Chapter 2

Summary

Chapter 1 highlights the distinct characteristics of auditory streaming, two of them being build-up of stream segregation and bistability. Build-up refers to the tendency of a stimulus to be perceived as segregated with an increase in exposure time. Bistability in a percept refers to an ambiguous stimulus that could be perceived as either one of two possible perceptual patterns. This chapter uses the canonical A-B-A_ tone triplet stimulus used by van Noorden (1975) to investigate the effect of directed attention on streaming build-up for stimulus within the bistable range of streaming using a combined EEG and psychophysics paradigm. A repeated stimulus sequence of the ABA_ triplets was diotically presented and listeners were asked to either attend to an integrated or segregated percept. Subjective psychophysical responses showed a significant effect of attention on the rate of streaming build-up. Additionally, EEG responses showed an increase in activity related to the segregated percept in the latter half of the stimulus presentation compared to the first half of the stimulus presentation indicating an effect of streaming build-up.

2.1 Introduction

The concepts of stream segregation, bistable stimuli as well as the build-up of stream segregation have been introduced in Chapter 1. It is known that the segregation of sounds into auditory streams depends on several factors; two of them being the frequency difference between sounds and the rate of sound presentation. For intermediate frequency separations and presentation rates, deemed to fall within the 'ambiguity region' (Moore and Gockel, 2002), the percept may often 'flip' between one and two streams. The property of a stimulus to elicit such ambiguous percepts is called bistability (Moore and Gockel, 2012). The various aspects of bistability have been highlighted in several recent studies (Denham & Winkler, 2006; Pressnitzer & Hupé, 2006; Winkler, Denham et al., 2012). Lastly, it has been noted that one of the distinct characteristics of stream segregation is the build-up of perceptual streaming with increasing exposure time to the stimulus (Anstis & Saida, 1985; Bregman, 1978; Carlyon et al., 2001; Hupé & Pressnitzer, 2012). Hupé and Pressnitzer (2012) further describe build-up as 'a combination of a systematic bias towards the one stream interpretation at stimulus onset and a longer duration of this first percept compared with subsequent one stream percepts'. This initial extended one-stream percept has been termed by them as the 'inertia of the first percept' (Hupé and Pressnitzer, 2012). The concept of a systematic bias towards a one-stream percept has been questioned in a few recent studies (Deike et al., 2012; Winkler et al., 2012) where they suggest that the initial percept may not always be an integrated, one-stream percept. Anstis and Saida (1985) have also suggested that there is a distinct long-term trend as observation time increases to perceive the stimulus as segregated. However, this view has been refuted (Denham and Winkler, 2006; Denham et al., 2008; Pressnitzer and Hupé, 2006), especially for bistable percepts (Deike et al., 2012; Winkler et al., 2012) where results indicate that after the first initial percept, the trend of percepts followed in streaming is ‘purely stochastic’ and there is no distinct trend seen towards a long-term segregated percept (Pressnitzer and Hupé, 2006).

Some studies have maintained that attention is crucial for build-up of streaming even when the process is mainly stimulus driven (Alain and Woods, 1997; Brochard et al., 1999; Carlyon et al., 2001) whereas others have suggested that primitive stream formation can occur without focussed attention (Bregman, 1990; Macken et al., 2003; Sussman and Winkler, 2001; Sussman et al., 2003; Sussman et al., 2007). The role of attention in streaming has been described in detail in Chapter 1.

The effect of attention on bistable stimulus sequences has been of interest over the past decade as the listeners’ percept of the stimulus can be manipulated without any physical manipulation of the stimulus parameters. In this way, one can focus on the perceptual as well as evoked responses based solely on the difference in attention (Gutschalk et al., 2005; Snyder et al., 2006; Thompson et al., 2011).

Using the P1-N1-P2 evoked complex (using EEG or MEG) to tones to study streaming and effects of attention in bistable stimuli has been increasingly used (Gutschalk et al., 2005; Snyder et al., 2006; Szalárdy et al., 2013) as it can track the on-going processing of the stimulus (Snyder et al., 2006). All of these experiments are based on the relationship noted between streaming and neural adaptation (Fishman et al., 2001). A detailed explanation of the how the P1-N1-P2 complex is used to study streaming has been given in section 1.2.6.3.

Gutschalk et al. (2005) used the P1m-N1m-P2m complex elicited by a repeating ABA_ tone triplet paradigm in an MEG study to further examine the neural bases of streaming within a bistable percept range (frequency separation of 4 or 6 semitones). In the first experiment, manipulating the acoustic parameters of the sequence by systematically increasing the frequency separation between the A and B tones altered the percept. In the second experiment, the stimulus parameters were held constant within the bistable region where listeners spontaneously switched between an integrated or a segregated percept**.** Listeners were instructed to report their percept (1 or 2 streams heard) via button press and were told to attend to either the A tones or B tones whenever they heard a segregated percept. In both cases of directed attention during the segregated percept, they found an enhanced response in the P1m-N1m wave following the B tone (see section 1.2.6.3 for a detailed explanation). Overall, they found a strong coupling between the streaming percept and auditory cortical activity in the absence of any stimulus changes and suggested percept-dependent modulation of the P1m and N1m responses (early sensory processes).

Snyder et al. (2006) used an evoked potentials (P1-N1-P2) paradigm in EEG to study the effect of attention on stream segregation using the ABA_ tone triplet paradigm. Their results suggested that the evoked potentials became larger for trials where the listeners were instructed to focus their attention on the auditory stimulus as opposed to trials where they were instructed to ignore the stimuli (i.e. passive listening). Also, the build-up process and amplitude of the evoked potentials was highly modulated by attention. However, they found that the EEG activity picked up automatic stream segregation of the sounds in the passive listening condition as well. Overall, they suggest that their experiment provides evidence for two mechanisms, one being the automatic segregation of sounds and the other being an attention dependant build-up of segregation that is heavily modulated by attention (Snyder et al., 2006). Based on an MMN study, Sussman et al. (2007) also suggested that although attention is not always needed for segregation, one cannot rule out that attention plays a role when the acoustics fall in the ambiguous range of stimuli.

The primary aim of the current experiment was to study the effect of attention on the build-up of stream segregation for a tone triplet within a bistable frequency range. As mentioned above, it has previously been studied using EEG for long durations of ABA_ tone triplets (Snyder et al., 2006). In the present experiment, an additional factor of a temporal shift was added to obtain a more objective measure of build-up. The use of the temporal shift has been described below.

The relation between auditory stream segregation and the perception of temporal relationships between sounds has been explored in earlier studies (Haywood and Roberts, 2013; Micheyl and Oxenham, 2010; Roberts et al., 2008; van Noorden, 1975; Vliegen et al., 1999). In general, in an on-going stimulus sequence of ABA_ tone triplets, a temporal shift of the B tone (the B tone in one of the triplets is delayed by a certain amount of time) is easier to detect when listeners rely on within-triplet timing judgements (Micheyl and Oxenham, 2010), which is possible when the sequence is heard as an integrated percept. It has been shown that stream segregation impedes performance on temporal shift detection tasks (Vliegen et al., 1999). Furthermore, the temporal shift should be maximally detected when the shift is early on in the stimulus sequence and when the listeners are instructed to hold on to an integrated percept.

Stream segregation build-up can also be affected to varying degrees by introducing an abrupt change in a stimulus sequence (Haywood & Roberts, 2013; Roberts et al., 2008). Several studies have suggested that changes in frequency, intensity, rhythm, lateralisation or attention could cause a possible reset of the streamed percept and have investigated this using either psychophysical paradigms (Cusack et al., 2004; Denham et al., 2008; Roberts et al., 2008; Thompson et al., 2011) as well as electrophysiological methods (Szalárdy et al., 2013; Thompson et al., 2011). A resetting of stream segregation suggests that in an on-going sequence of stream segregation where the percept starts from a one-stream integrated percept and builds up to a two-stream percept, a sudden change via any of the factors mentioned above (frequency, intensity, temporal pattern, lateralisation or attention) could break the build-up of segregation and cause the percept to be reset to the initial, default one-stream percept (Anstis and Saida, 1985; Rogers and Bregman, 1998; Roberts et al., 2008). On the other hand, studies have also shown that a change (or deviant) in an on-going sequence may be perceived in a deviant-detection task (for example, detecting a temporal shift in an on-going ABA_ tone sequence) but may not cause a resetting effect in the perceptual build-up of streaming (Denham et al., 2008; Haywood and Roberts, 2013; Roberts et al., 2008; Thompson et al., 2011).

In the current experiment, this concept of temporal shift detection and its relation to the streaming percept is incorporated as a measure of build-up in an on-going ABA_ tone triplet stimulus sequence. The experiment consists of two main attention conditions, one that promotes integration and one that promotes segregation of the tones into separate streams. Based on the studies described above, the primary hypothesis for this experiment was that the temporal shift in the attention conditions promoting integration would cause a larger change in the EEG than in the attention conditions promoting segregation. The second hypothesis was related to the position of the temporal shift in the stimulus. The experiment had two positions where the deviants were placed in the stimulus sequence, one early on and one later in the stimulus. The second hypothesis of the experiment was that the early deviants would cause a larger amplitude change in the EEG than the later deviants, indicating build-up of segregation, as presumably, the initial deviant would still fall under the integrated percept for both conditions. It was however hypothesised that the deviant would be maximally detected in the EEG for the integrated-attention condition with the early-placed temporal shift. Lastly, it was hypothesised that the subjective responses would possibly show an effect of attention on streaming build-up and might show a perceptual resetting of the subjective response due to the temporal shift.

2.1 Methods

2.1.1 Participants

Twelve adults (25–35 years of age, 5 male) with normal hearing (audiometric thresholds below 20 dBHL for octave frequencies between 250 Hz and 8KHz) and reportedly no history of neurological disorders participated in the experiment. Informed consent was obtained after the procedures of the experiment were explained to them. They received payment for their participation. The University College London Ethics Committee approved the experiment.

2.1.2 Stimuli and conditions

Repeated ABA_ tone triplets were used where A and B are pure tones of frequencies 1260 Hz and 1000 Hz respectively (4-semitones apart). Each tone was 100 ms in duration including 10 ms onset and offset cosine ramps. For a given A-B-A_ triplet, the within-triplet silent duration was 50 ms and the inter-triplet silent duration was 250 ms. Each stimulus sequence was made up of 15 repetitions of these triplets (15 tone triplets in total), with a total duration of 9.75 seconds (see Figure 2.1 for a schematic of the stimulus parameters). The stimulus was generated in MATLAB (MathWorks Inc. Natick, MA, USA) with 16-bit resolution and at a sampling frequency of 44.1 kHz. Stimuli were presented through the Neuroscan Stim2 software diotically through Etymotic Research ER-2 insert transducers (Etymotic Research, Elk Grove Village, IL, USA) at a comfortable listening level of 70 dB sound pressure level (SPL) through an external sound box.

There were two attention condition groups, which had 3 sub-conditions each. In the first attention condition group, the listener had to direct attention towards a galloping percept involving all the tones (to promote integration). The second attention condition group required the listener to direct attention towards the B tones in particular (to promote a segregated percept). Within each attention group, the first sub-condition was a control condition with no temporal shift, the 2nd sub-condition had a temporal shift of 40 ms early in the stimulus sequence (shift in the 4th triplet out of 15 triplets in the 9.75 second stimulus sequence) and the 3rd sub-condition had a temporal shift later in the stimulus sequence (shift in the 11th triplet out of 15 triplets in the 9.75 second stimulus sequence) (see Figure 2.1).

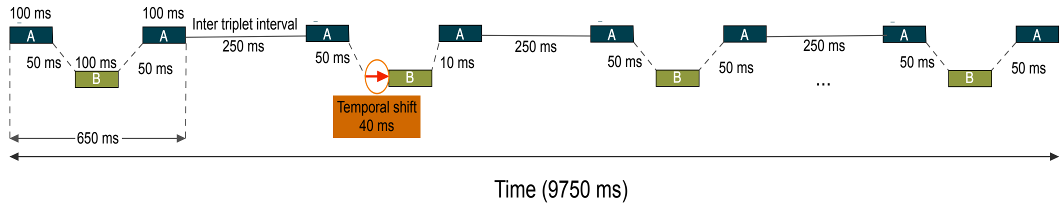


Figure 2.1: *Schematic representation of a typical repeated tone triplet sequence used in the stimulus with a temporal shift on the B tone.*

2.1.3 Procedure

Participants were seated in a sound-attenuated room in a comfortable chair in front of a computer monitor that displayed the visual prompts indicating the start and end of each stimulus. Before the start of the 9.75 seconds stimulus, a green circle was displayed on the screen to indicate the beginning of the trial. This was followed by a 9.75-s stimulus sequence of 15 repeated ABA_ tone triplets.

Before the test session, the listeners were given demonstrations of what was meant by a one-stream and a two-stream percept with stimuli, which were easily perceived as one or the other. For example, a simple demonstration of a 2 stream segregated percept involved listening repeated tone triplets with a 12-semitone difference at fast rates (easily segregated). They were, however, not told about what pattern of percept to expect throughout the test sequence to avoid biasing their responses.

Participants responded overtly via the Neuroscan response box by button-press to indicate if they heard 1 or 2 streams throughout the presentation of the stimulus. The participants were instructed to keep their arms still and only respond with their thumbs using minimal movement. Continuous EEG was recorded throughout the test session.

A previous control EEG experiment carried out on 10 listeners with similar repeated tone triplets indicated that there was no significant interference of muscle artefacts due to button-press responses on the concurrent EEG recording. This was tested by assessing if there was any difference in the N1-P2 latency/amplitude with or without button-press response to an on-going ABA_ streaming stimulus with a four-semitone separation (within the bistable range). Listeners in this experiment were not instructed to attend to any particular percept (neither integrated nor segregated).

For each condition block, participants were instructed to direct their attention either towards the galloping non-segregated percept throughout the test block or to attend to the B tones in the triplets. The participants were not aware of the temporal shift in the four conditions.

2.1.4 EEG recording

The EEG activity was recorded using a NuAmps amplifier with the Acquire option of Scan 4.2 software (Neuroscan, 2003) . The signal was digitally sampled at an A/D rate of 1000 Hz (16-bit resolution). Listeners were fitted with an electrode cap fitted with 32 silver/silver chloride scalp ring electrodes positioned in an electrode ‘Easy Cap’ (Falk Minow Services, Herrsching-Breitbrunn, Germany) for 32/40 channels. The electrode positions from which EEG was recorded were FP1, FP2, Fz, F3, F4, F7, F8, FCZ, FC3, FC4, FT7, FT8, CZ, C3, C4, TP7, TP8, CPZ, CP3, CP4, PZ, P3, P4, OZ, O1, O2, FT9 and FT10. Eye movements were recorded using two horizontal and two vertical electrodes: the vertical electrooculogram (EOG) was measured between one electrode placed above and another below the left eye and the horizontal EOG was measured between electrodes placed lateral to the outer canthi of both eyes. Electrodes were attached using Quik-Gel Conductive Gel and low impedances (<10kΩ) were ensured. Each sub-condition consisted of 100 repetitions (i.e. 600 trials in total). The stimulus, visual prompts and button press responses had unique trigger codes enabling acquisition of individual time stamps for each event per trial.

2.2 Results

2.2.1 Behavioural data

Response data from each trial were divided into fifteen 650-ms time bins corresponding to each triplet (Figure 2.2). Within each time bin, the number of streams heard by the listener was averaged over the 100 trials for each listener per condition. Data was then averaged across the 15 listeners as a function of time within each time bin. These data are displayed in Figure 2.2.

A two-way repeated measures ANOVA was conducted on the behavioural response data (Condition X Time interval: 15 time bins of 650 ms each), confirmed significant main effects of condition, $F(5,50) = 8.13$, $p<0.001$; and time interval, $F(14,140)=37.43$, $p<0.001$. A two way interaction between condition and time interval was also significant, $F(70,700) = 3.99$, $p<0.001$. Post hoc tests indicate that the three conditions where the listeners were instructed to focus on an integrated percept differ significantly ($p<0.005$) from the three conditions where the listeners were instructed to focus in on the B tones (which promoted segregation). The conditions aimed at promoting segregation show a faster build-up of stream segregation (seen by the lines connecting the square symbols in Figure 2.2).

There was no apparent effect of the temporal shift on the perceptual build-up of streaming (as noted by the blue (early temporal shift) and green lines (late temporal shift) for both attention conditions). There was no resetting toward an integrated percept seen in any of these conditions. These results showing a lack of resetting are in line with other studies investigating build-up and resetting effects of abrupt changes in an on-going stimulus sequence (Denham et al., 2008; Haywood and Roberts, 2013; Roberts et al., 2008; Thompson et al., 2011).

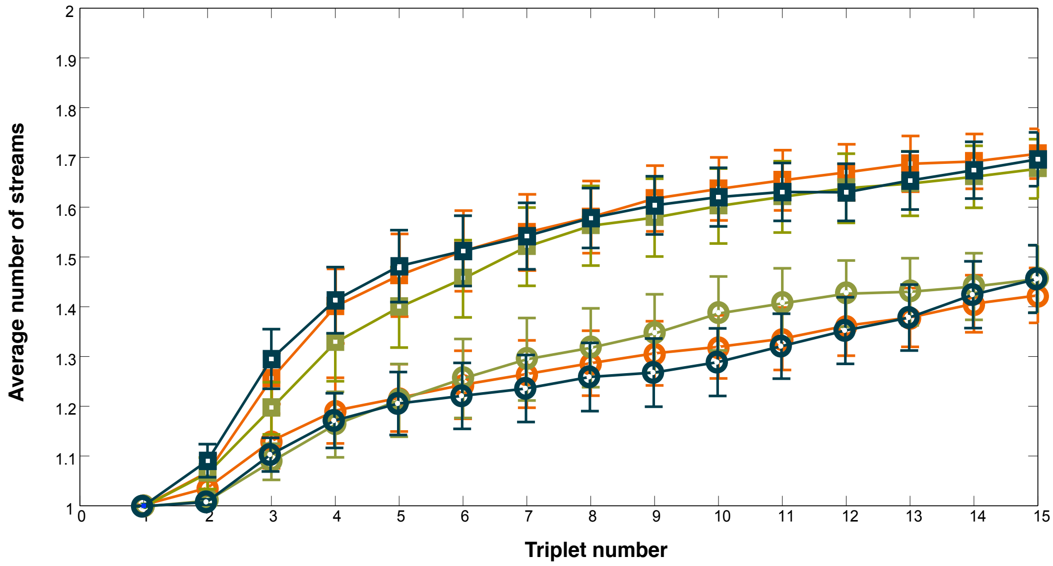


Figure 2.2: *Behavioural data results averaged from 15 listeners. The orange lines indicate the control-no shift conditions, the green lines indicate the early-temporal shift conditions and the blue lines indicate the late-temporal shift conditions. The square data points (upper three data lines) indicate the conditions where the listeners were instructed to focus on the B tones to promote segregation. The circle data points indicate the conditions where the listeners were instructed to focus on an integrated percept.*

2.2.2 EEG data

2.2.2.1 Pre-processing, averaging across participants and segregation of data

EEG pre-processing, epoching and averaging was carried out using the EEGLAB toolbox (Delorme & Makeig, 2004). Data was filtered using a zero phase shift band pass filter from 0.1 Hz to 30 Hz. Baseline was corrected to -100 ms followed by artefact rejection at +/- 150 microvolts. Data was averaged across the 100 trials per condition.

From the continuous EEG data, epochs of 9850 ms were extracted, starting 100 ms before the initial tone triplet onset. The averaged data across the 15 participants is shown in Figures 2.3, 2.4 and 2.5.

2.2.2.2 Effect of temporal shift

Our initial hypothesis aimed to see if there was a difference in the P1-N1-P2 waveform elicited by the temporally shifted tone in the conditions where the B tone was shifted by 40 ms. A large P1-N1-P2 wave was expected to be seen corresponding to the deviant in all four conditions. However, according to the hypothesis, an effect of directed attention would be seen as a difference in the amplitude of the P1 and N1 waves between conditions where the physical stimulus was the same but the listeners were instructed to either focus on an integrated or segregated percept. It was hypothesised that an increase in amplitude would be noted for conditions where listeners were instructed to focus on an integrated percept.

The averaged EEG waveforms shown in Figure 2.3 and 2.4 show the waveform with the temporal shift in the early shift condition and late shift condition. Based on bootstrap based t-intervals (two-tailed, $p < 0.005$) calculated for the whole source waveform), a significant increase in amplitude in attention conditions in the early shift condition on the P1-N1-P2 complex elicited by the deviant B tone (larger amplitude for the condition was observed where listeners were instructed to attend to the integrated percept). However, there is no significant difference between the attention conditions in the late shift condition. As is shown in Figures 2.4 and 2.5, the deviant tone elicited a large response in all 4 conditions; however, the amplitude of the response was largest in the condition where the participants had to attend to an integrated percept and the temporal shift was early on in the stimulus.

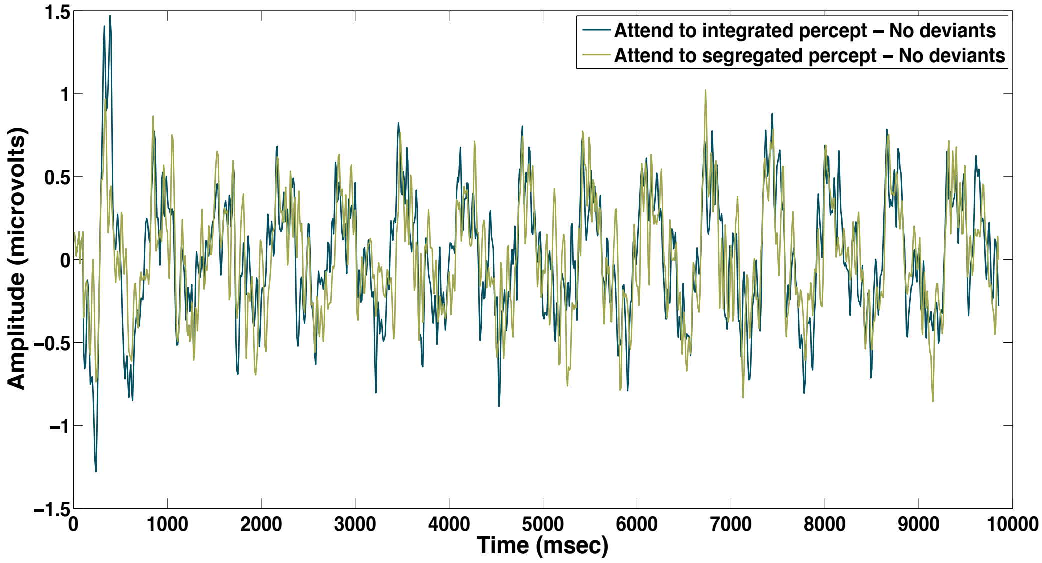


Figure 2.3: *Waveform of conditions with no temporal shifts for the two different attention conditions. Dark blue lines indicate average waveform for the no-shift condition where listeners were instructed to hold on to an integrated percept. Green lines indicate average waveform for the no-shift condition where listeners were instructed to hold on to a segregated percept.*

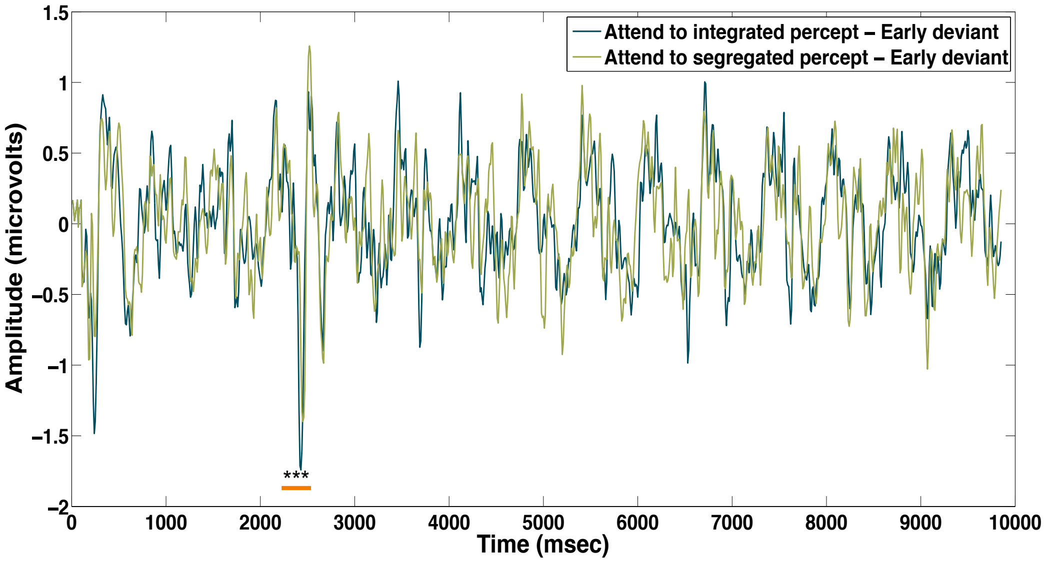


Figure 2.4: *Waveform of early temporal-shift conditions for the two different attention conditions. Dark blue lines indicate average waveform for the early-shift condition where listeners were instructed to hold on to an integrated percept. Green lines indicate average waveform for the early-shift condition where listeners were instructed to hold on to a segregated percept. A significant difference was noted in the activation for the temporally shifted tone (indicated by orange line) (p<0.005) where the amplitude was greater for the condition where listeners were instructed to attend to the integrated percept*

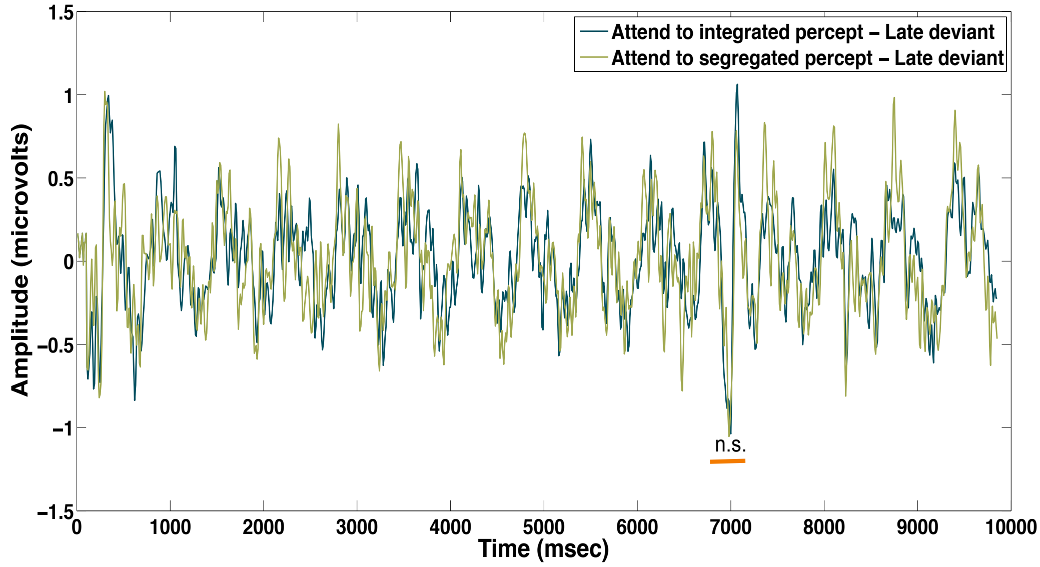


Figure 2.5: *Waveform of late temporal-shift conditions for the two different attention conditions. Dark blue lines indicate average waveform for the late-shift condition where listeners were instructed to hold on to an integrated percept. Green lines indicate average waveform for the late-shift condition where listeners were instructed to hold on to a segregated percept. No significant difference was noted in the activation for the temporally shifted tone (indicated by orange line).*

2.2.2.3 Effect of attention results

As noted from Figures 2.3, 2.4 and 2.5, a large degree of overlap of the individual components of each of the tones is present because of the inter-tone interval being relatively small. Instead of analysing the data in terms of peak amplitudes of the P1-N1-P2 peaks elicited by each tone, the data was analysed in terms of the rates of the different components in the stimulus. The two main frequencies that were tested were the frequency of the occurrence of every tone (the occurrence of either A or B tones) and the frequency of occurrence of only the B tones. In order to ensure that the timing between every tone of the stimulus was equal, the EEG signal to the exact length of each of the tone triplets plus 50 ms of silence was extracted and concatenated. This ensured that the activation for each of the tones was separated by a 50 ms time difference. The time points for the start and end of each tone triplets were marked and extracted and the EEG for each of these 15 tone triplets was then concatenated.

In this new concatenated sequence, the time between the onsets of each of the tones was 150 ms (100 ms tone length plus 50 ms silence), which resulted in a rate of 6.6 Hz. Now, if the rate of the occurrence of the B tones in the new concatenated sequence was calculated, the time between the onsets of the B tones would be 450 ms (3x100 ms tone lengths plus 3x50 ms silences), which would result in the rate of 2.2 Hz. The EEG was then averaged across all three sub-conditions (no-deviant, early-deviant and late deviant) for each attention condition. A Fast Fourier Transform was then carried out on this averaged concatenated sequence of the EEG only related to the tones for each participant. Figure 2.6 shows the mean of the peaks at the two frequencies across all participants. The upper panel of Figure 2.6 shows the power spectrum for the initial half of the stimulus whereas the lower panel of Figure 2.6 shows the power spectrum for the latter half of the stimulus. It was found that for the integrated percept, the amplitudes of the peaks did not change significantly. However, for the segregated attention condition, it was found that the peak related to the frequency of the B tones increased significantly in the second half of the stimulus as compared to the amplitude in the first half of the stimulus. A repeated measures ANOVA with factors Section (early v/s late), Attention condition (integrated v/s segregated) and Frequency (2.2 v/s 6.6) was conducted on the EEG frequency data. Results confirmed significant main effects of section, $F(1,10)=5.13$, $p<0.01$; attention, $F(1,10)=4.6$, $p<0.01$ and frequency, $F(1,10)=8.52$, $p<0.01$ with a significant interaction between section, attention and frequency: $F(1,10)=16.615$, $p<0.01$. Post hoc tests indicated a significant difference ($p<0.01$) between the amplitude at 2.2 Hz as well as 6.6 Hz (frequency of the B tones and frequency of every tone respectively) for the segregated percept across the early and late part of the stimulus.

These results can be interpreted as follows:

1) The EEG activity at the rate of the B tones (2.2 Hz) increases in the later half of the test sequence as compared to the earlier half of the stimulus for the attention conditions where the listeners were encouraged to segregate the streams compared to the attention conditions where the listeners were encouraged to hold on to an integrated percept.
2) The activity at the rate of all the tones (where each tone is heard) decreased in the later half of the stimulus as compared to the earlier half of the stimulus for the attention conditions where the listeners were encouraged to segregate the streams.

Hence, the effect of attention on build-up of segregation was indicated by the increase in the amplitude of the evoked response to the slower rate in the attention condition where the listeners were instructed to segregate. In the case of the conditions where the listeners were encouraged to perceive an integrated percept, there was a smaller difference in activity across the early and late halves of the stimulus at the rate of the B tones.

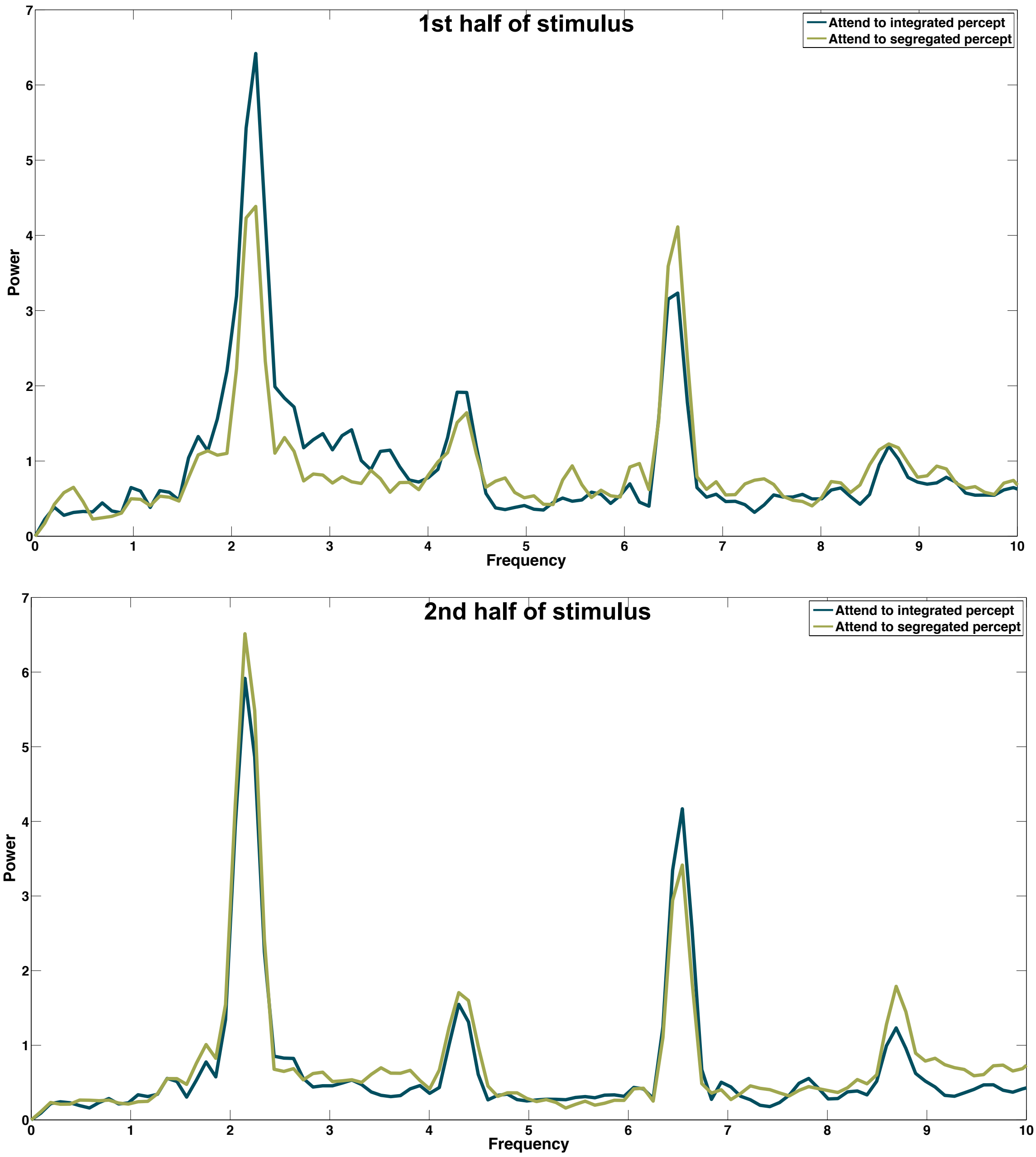


Figure 2.6: *Upper panel: Frequency spectrum for the two attention conditions (indicated by the blue and green lines) for the frequency of the B tones (2.2 Hz) and the frequency of occurrence of every tone (6.6 Hz) for the first 4.9 seconds of the stimulus. Lower panel: Frequency spectrum for the two attention conditions (indicated by the blue and green lines) for the frequency of the B tones (2.2 Hz) and the frequency of occurrence of every tone (6.6 Hz) for the later 4.9 seconds of the stimulus.*

2.3 Discussion

This experiment aimed at investigating the way in which attention affects the build-up of stream segregation. The experimental hypothesis was that the behavioural as well as EEG data would show an increase in build-up of segregation for conditions where listeners were instructed to attend to a segregated percept (try to segregate the B tones from the A tones) as opposed to an integrated percept (where participants were instructed to hold on to an integrated ABA_ percept). In the behavioural data, for the conditions where the participants were instructed to attend to the B tones, a much faster build-up was observed as compared to the conditions where the listeners were instructed to attend to the one stream 'galloping' percept. There has been an on-going debate regarding the role of attention in build-up of segregation (Moore and Gockel, 2012). Carlyon et al. (2001) have demonstrated that build-up of segregation depends on attention. They stress that even though they cannot conclude that unattended sequences do not split into separate streams, they can demonstrate that attention has a large effect on build-up and they argue that attention is crucial for the process of streaming sounds into separate streams rather than simply a modulating factor for the output of the streaming process (Carlyon et al., 2001).

The EEG data for the six conditions showed a smaller, yet significant effect of attention on the build-up of streaming. Two aspects of attention were studied: one being the difference in the amplitude of the responses evoked by the deviant tone and the other being a measure of the increase in activity at the repetition rate (frequency) of the segregated B tones. As hypothesised, the deviant was detected more (as indicated by an increase in amplitude of the P1-N1-P2 complex) in the 'early' condition than the later condition. Even within the early conditions, it was detected maximally for the attention condition where the listeners were supposed to hold on to an integrated percept, as hypothesised. This could be due to the fact that in the integrated percept, since each tone is perceived, the perceived inter tone interval is small (50 ms). A 40 ms shift in an on-going rhythm with only a 50 ms inter-tone interval is easily detectable. However, in the case of a segregated percept, the inter-tone interval becomes much larger (550 ms between two B tones) in which case, a 40 ms shift in tones is not as salient. In the later condition, there was no difference between the attention conditions for the amplitude of the response evoked by the 'late' deviant.

A significant increase in the rate of the segregated B-tone stream towards the latter half of the stimulus for the condition was also noted where the listeners were instructed to attend to the B tones to enhance segregation. It must be noted that in the behavioural as well as the EEG responses, the percept tended towards a two-stream percept even when listeners are instructed to hold on to an integrated percept. However, the amount of streaming build-up was more rapid and significantly larger for the attention conditions where stream segregation was facilitated. This is in line with the results from (Carlyon et al., 2001; Sussman et al., 2007) which indicate that although attention is not essential for build-up of segregation, it does play a critical role in the rate and extent of segregation. There have been several recent studies suggesting that auditory attention can modulate EEG responses in a manner that is consistent even at single-trial level (Kerlin et al., 2010; Choi et al., 2014). Furthermore, Woldorff et al. (1993) have also shown that auditory attention can exert selective control over sensory input as early as 20 ms post stimulus onset.

The final aspect that was investigated in this data was the level of resetting of stream segregation due to the inclusion of the deviant.

The paradigm of the current experiment was very similar to that used in a psychoacoustic study by Thompson et al. (2011). In their first experiment, they used a 25 pure tone triplet sequence with an early or late deviant tone for 4 and 8 semitone frequency differences. The participants made subjective and objective responses separately for the same stimuli; subjective responses involved making perceptual judgments throughout the stimulus to decide if they heard either one or two streams whereas the objective task required the listeners to detect the temporal shift via button press. Their results showed no evidence of resetting of streaming (Thompson et al., 2011). The same lack of resetting from a segregated 2-stream percept to a default 1-stream percept in the subjective response was seen in our data. However, the temporal shift in the stimulus was very clearly detected in the EEG activity in our current experiment as seen by the change in the P1 and N1 wave amplitudes. This could suggest that the shift, although detected, does not cause a simultaneous reduction in the reported subjective streaming percept. Post testing, when listeners were asked if they had noticed the temporal shift, only 3 out of 15 listeners had been able to detect the temporal shift. However, even those 3 listeners did not show any active resetting in their psychophysics data. This result related to the conclusions drawn by Roberts et al. (2008), Haywood and Roberts (2013) and Denham et al. (2008) where they stated that abrupt acoustic changes may affect temporal judgment measures but may not necessarily cause any ‘reduction in the reported extent of segregation’ (Roberts et al., 2008).

Overall, the results of this study indicate that the effects of attention and build-up can be seen in the behavioural as well as the EEG responses for bistable stimuli. The results of this chapter lay the groundwork for the study described in Chapter 3, which describes the use of a novel stimulus, based on the octave illusion (Deutsch, 1974) used to study stream segregation for sequential as well as synchronous sounds using concurrent EEG and psychophysics techniques.

Chapter 3

Summary

This chapter introduces the use of a perceptually multistable auditory stimulus, similar to the one used to elicit Deutsch's 'Octave Illusion', which involves both sequential and concurrent sound segregation. This stimulus has been revisited in the following set of studies as it provides a new potential insight into understanding the role of attention in stream segregation. In both the studies described in this chapter, the basic structure of the stimulus remains similar to that of Deutsch's octave illusion. Each ear of the listener was presented with a sequence of alternating pure tones of low and high frequencies. The same sequence was presented to each ear, but in opposite phase, such that the sequence in the left ear could be a High-Low-High… pattern whereas the sequence in the right ear was a Low-High-Low… pattern. The illusion reported by Deutsch (1974) is that participants hear an alternating pattern of low and high tones, with all the low tones lateralized to one side and all the high tones lateralized to the other side. In the current set of studies, the stimulus sequence was preceded by a priming sequence of tones that were either all low or all high in frequency and were presented either to the left or right ear. Listeners were cued to focus on a particular frequency (high or low) and side (right or left), as indicated by the priming sequence

of tones. By instructing participants to listen to a particular frequency and side, two different percepts for the same stimulus were possible, thus allowing us to study the neural correlates of streaming and selective attention. Two experiments were carried out to test the hypotheses that 1) stream segregation is key to the perception of the illusion and 2) the illusion can be successfully manipulated via directed attention. Firstly, psychophysical measures were used to establish if the illusory percept elicited by this particular stimulus configuration depended on stream segregation. Psychophysical results indicated that streaming played a role in the way this stimulus was processed. Next, a concurrent EEG and psychophysics paradigm was used to objectively establish if the effect of attention was mirrored in the EEG recordings.

3.1 Introduction

The previous chapter used a well-established sequential streaming stimulus, the ABA_ tone triplet paradigm, to study the effect of directed attention on the perception of a multistable stimulus and investigated whether the effects of attention can be observed in the resultant EEG signal. The results indicated that there was, indeed, an effect of directed attention in the EEG signals and the subjective psychophysical results also showed a distinct effect of attention on the outcome of auditory streaming.

However, as discussed in Chapter 1, more often than not, sounds in the environment occur both concurrently as well as sequentially in time. Auditory stream segregation has been studied in various psychophysical and physiological experiments using sequential and concurrent sounds. As discussed in chapter 1, the acoustical properties that influence streaming of stable and multistable stimuli are relatively well known and have been explored using several behavioural as well as electrophysiological methods (Alain, 2007; Denham et al., 2010; Elhilali et al., 2009; Gutschalk et al., 2005; Micheyl et al., 2007, 2013a, 2013b; Pressnitzer & Hupé, 2006; Shamma & Micheyl, 2010). Furthermore, the role of attention, expectation, prior exposure and other “top-down” influences have been found to play a key role in how complex auditory scenes, especially ones with perceptually ambiguous stimuli, are perceived and processed (Carlyon et al., 2001; Cusack et al., 2004; Elhilali & Shamma, 2008; Moore & Gockel, 2012; Winkler et al., 2012). It must be noted that perceptual ambiguities are relatively rare in natural settings mainly due to several other cues helping listeners to disambiguate stimuli. However, perceptually ambiguous or multistable stimuli can be useful in dissociating the neural responses to physical stimuli from the correlates of perception (Leopold and Logothetis, 1999; Schwartz et al., 2012). Furthermore, although sounds in the natural environment occur both synchronously as well as sequentially, not many studies have been conducted in the field of stream segregation to study how these two aspects of segregation interact. In general, with a few exceptions (Darwin et al., 1995; Shinn-Cunningham et al., 2007), the perceptual organization of sequential and concurrent sounds has been studied separately and independently.

The experiments described in this chapter address the interaction between sequential and concurrent sound segregation using a new stimulus paradigm similar to the one used to elicit Deutsch's "octave illusion" (Deutsch, 1974) (Figure 3.1). The illusion reported by Deutsch is that participants hear an alternating pattern of low and high tones, with all the low tones lateralized to one side and all the high tones lateralized to the other side, whereas the actual stimulus has alternating low and high tones in both ears.

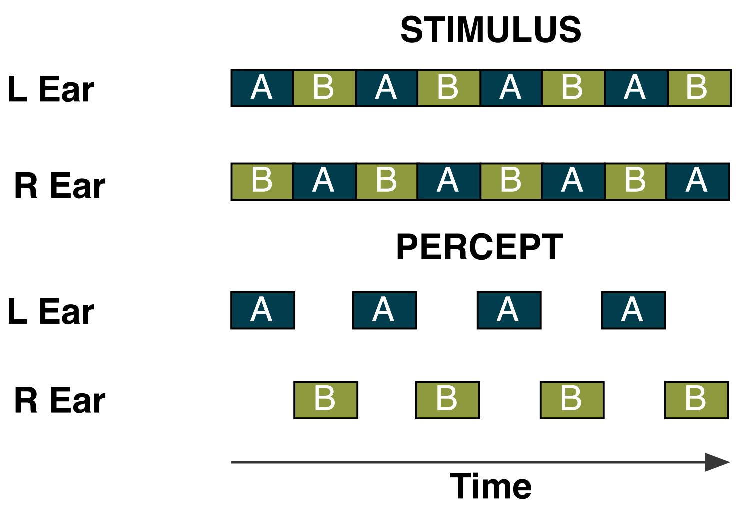


Figure 3.1: *The stimulus pattern used in the original experiment of Deutsch (1974) describing the octave illusion, together with the percept most commonly obtained. Blue boxes indicate tones at 800 Hz, and green boxes indicate tones at 400 Hz.*

In has been noted that all participants do not perceive the illusion in the same fashion (the pattern may differ from high tones in right ear and low tones in left ear to the opposite pattern) and that the pattern depends on the length of stimulus presentation, i.e. the illusion does not occur for very short presentations of the stimulus (Christensen & Gregory, 1977; Deutsch, 1978). Given that this stimulus can be perceived in multiple ways, it is reasonable to claim that this stimulus sequence could indeed be multistable (Deutsch & Gregory, 1978; Chambers et al., 2002; Brancucci et al., 2011; Brancucci et al., 2014). Although studies have noted multiple percepts for these stimuli, none have investigated whether instructions or priming (either verbal or auditory priming) can be used to alter the percept. For example, if listeners are instructed to listen for the low tones in the left ear, can they do this successfully, or does their percept revert to hearing the low tones in the right ear, as reported in the original illusion? While a few studies have probed the neural representation of these stimuli (Lamminmäki and Hari, 2000; Brancucci et al., 2012) as discussed in Chapter 1, none of these studies attempted to actively manipulate the percept while simultaneously recording neural responses. Also, the majority of studies carried out using this illusory stimulus do not use objective measures to record the participants' percept.

The following two experiments test the hypotheses that stream segregation is key to the perception of the illusion and that its perception and neural correlates can be manipulated via selective attention. In order to ensure the listeners' attention was manipulated adequately, an objective deviant detection paradigm was used where listeners had to detect a target amplitude deviant within the attended streams while ignoring distractor deviants in other streams.

3.2 Methods

3.2.1 Participants

Fifteen participants (eleven female and four male, aged 20-29 years) participated in the first experiment, which involved only psychophysics. Ten participants (four female and six male, aged 20-29 years) participated in the second experiment, which involved simultaneous behavioural and EEG measurements. All participants tested were naïve listeners and there was no overlap of participants between the studies. All participants had normal hearing, defined as audiometric hearing thresholds no more than 15 dB Hearing Level (HL) at octave frequencies from 250 Hz to 4 kHz, with no history of hearing or neurological disorders. Participants provided written informed consent and were compensated for their participation. Experiment 1 was carried out at University College London and Experiment 2 was carried out at the University of Maryland. The University College London Ethics Committee and the University of Maryland Institutional Review Board approved the procedures for the experiments respectively.

3.2.2 Experiment 1: Stimuli and procedure

In the first experiment, alternating sequences of low and high tones were presented to each ear in opposite phase, such that the sequence in the left ear could be a High-Low-High… pattern while the sequence in the right ear could be a Low-High-Low… pattern (see Figure 3.2). Participants were cued to attend to a particular ear (R or L) and frequency (termed Hi or Lo), as indicated by a priming sequence of three pure tones that were presented either to the left or right ear and were either all low or all high in frequency (i.e., RLo, RHi, LLo, or LHi). All tones were 100 ms in duration, including 10-ms raised-cosine onset and offset ramps. All tones were presented at 70 dBHL. Within the priming and the main sequence, the tones were separated from each other by a 50-ms silent period. The silent period between the priming sequence and the test sequence was 500 ms. The sequences were generated in MATLAB (MathWorks Inc. Natick, MA, USA) and were presented at a sampling rate of 44.1 kHz. The experiment was presented using the Psychophysics Toolbox extension in MATLAB (Brainard, 1997; Pelli, 1997) through Sennheiser HD 215 headphones.

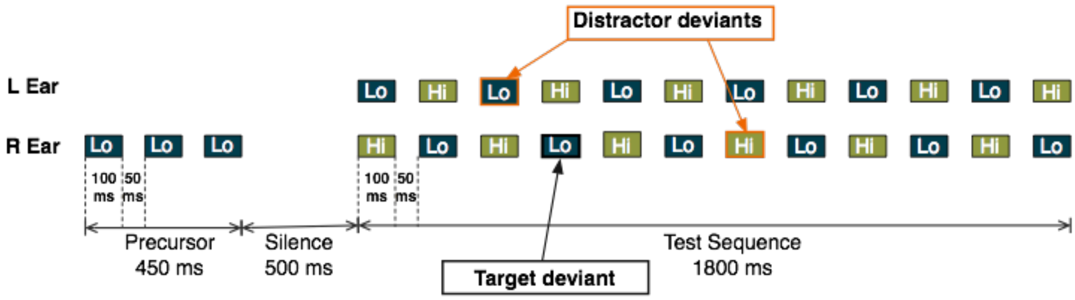


Figure 3.2: *Schematic representation of the stimuli. Blue and green boxes indicate pure tones of 1000 and 3000 Hz respectively. Each ear receives an alternating sequence of Hi-Lo tones. The example trial shown in the figure has a precursor sequence of low frequency tones in the right ear indicating the attended stream. The amplitude deviant in the Right-Low tones thus becomes the target deviant among the other distractor deviants.*

Each ear of the listeners was presented with alternating, opposing sequences of 12 pure tones per trial – six high and six low tones in each ear (see Figure 3.2). Each of the four tone streams (RLo, RHi, LLo, and LHi) could have one deviant tone (amplitude increase by 7 dB on one of the tones) that occurred either early, mid or late in the particular stream. Each stream had a randomized arrangement of the location of the targets and distractor deviants. It was ensured that an equal number of early, mid and late deviants were present across the test blocks. Depending on the priming sequence, the deviant in the primed stream was the target deviant, and the deviants in the other streams were termed distractor deviants. An example trial is shown in Figure 3.2, where the priming sequence is for the right ear and low tones (RLo), so the target is the deviant in the RLo stream and the distractors are deviants in any of the other streams (as indicated in Figure 3.2). The participants were required to detect the target deviant while ignoring all other distractor deviants. They responded via button press at the end of each trial to indicate whether a target deviant had been presented. The target deviants were present in 50% of the trials whereas distractor deviants were present in every trial. All the deviants used in the test sessions were 7 dB higher than the other tones in the sequence, based on listeners achieving a sensitivity index (d') of 1.0 or higher in pilot experiments with that increment level.

Four different frequency separations between the high and low tones were used: 1, 6, 15 and 20 semitones. Within a block, the order of presentation of trials was randomized for the four frequency separations and the four probe types. The frequency of the low tone was fixed at 1000 Hz while the frequency of the high tone varied between trials. Participants received visual feedback at the end of each trial.

Each participant undertook an initial session with 5 repetitions of the test sequence without any priming tone sequence at the maximum frequency separation of 20 semitones. For each trial, their unbiased percept (i.e., when they were not provided with instructions on what to attend to within the sound sequences) was noted. For this, the participants were asked to simply listen to the sound sequence and report what they heard. Next, they were presented with all the frequency separations (1, 6, 15 and 20 semitones) with the different priming sequences (high and low frequency priming tones in the right and left ear) to check if the listeners were able to manipulate their percept. For example, the participants were primed to LHi tones and their percept at the end of the test sequence was noted. Following this block, they carried out one practice block of the deviant detection task. For the actual test conditions, each participant completed 20 blocks of the task. Each block consisted of 96 trials. For each frequency separation, there were three deviant and three non-deviant trials per block. The order of all trials was fully randomized. Each block took approximately 10 minutes, depending on the participants' response time. The testing was broken up into two sessions of approximately two hours each.

3.2.3 Experiment 2: Stimuli and procedure

The second experiment combined EEG and psychophysical measurements to investigate the perception and neural representation for a stimulus similar to that used in Experiment 1. The primary difference was that the frequencies of the low and high tones were fixed at 1000 and 3000 Hz, respectively. As in experiment 1, each ear was presented with an alternating sequence of 12 pure tones per trial (see Figure 3.2). One amplitude deviant was placed on each of at least three of the four types of tones (RLo, RHi, LLo, LHi) either at the start, middle or at the end of the sequence. The sequences were generated in MATLAB (MathWorks Inc. Natick, MA, USA) and digitized at 44.1 kHz. The stimuli were presented using E-prime (Psychology Software Tools, Inc. Sharpsburg, PA, USA) through Etymotic Research ER-2 insert transducers (Etymotic Research, Elk Grove Village, IL, USA) in a sound treated room. Depending on the priming sequence, one of the deviants would be the target deviant and other three would be distractor deviants for that particular trial. The trials were designed in a way that the target deviants were present in 50% of the trials whereas the distractor deviants were present in every trial. The participants were required to detect the amplitude deviants in the stream of sounds that they were cued to (target deviants) and respond via button press at the end of each trial. Feedback was given at the end of each trial. Each listener was presented with 160 trials per priming condition during the test session.

EEG was acquired continuously using a 64-channel BrainVision system consisting of a Brain-Vision™ recorder (Version 1.01b) and a Brain-Vision professional BrainAmp™ integrated amplifier system (Brain Products GmbH, Germany). The signal was digitally sampled at an A/D rate of 1000 Hz (32-bit resolution). Participants were fitted with an electrode cap fitted with 64 silver/silver chloride scalp electrodes positioned in an electrode 'Easy Cap' (Falk Minow Services, Herrsching-Breitbrunn, Germany). Electrode impedance was monitored and maintained at a minimum (typically below 5 kΩ).

3.2.4 EEG Analysis

EEG pre-processing, epoching and averaging was carried out using the EEGLAB toolbox (Delorme and Makeig, 2004). Data was down-sampled and then filtered using a zero-phase-shift bandpass filter from 0.1 Hz to 30 Hz. Baseline was corrected to -500 ms, followed by artefact rejection at +/-150 microvolts. Independent component analysis (ICA) was used to remove artefacts related to eye movements and blinks.

The EEG signal for each attention condition (RLo, RHi, LLo, and LHi) was separated into epochs 2850 ms long including a 100 ms baseline. These were then grouped separately for correct trials (where the target deviants were correctly detected) and for incorrect trials (targets were either not detected or with false positive behavioural results). As most of the participants had d' values greater than 1.0, there were more correct epochs than incorrect epochs. Hence, for the second half of the analysis between correct and incorrect trials, a random subset of the correct trials was chosen to equal the number of incorrect trials in that condition.

The EEG activity was averaged individually for each of the four primed attention conditions: attend to RLo, RHi, LLo, and LHi (separately for correct and incorrect trials). Next, the responses were averaged with each pair of conditions that involved attention to tones that were presented synchronously. For example, for priming conditions of RLo and LHi, the evoked response waveform would show the same effect of attention, as the RLo and LHi tones are synchronous. In other words, the responses to the RLo and LHi conditions were averaged, as were the responses to the LLo and RHi conditions. Finally, the responses to the two pairs of conditions (RLo-LHi and LLo-RHi) were subtracted from each other in order to cancel out the common (in phase) 6-Hz activity (as the tone presentation rate in each ear was 6 Hz) and hence to potentially enhance the relative level of the 3-Hz activity (due to attention to alternate tones). Spectral analysis using a short-time Fourier transform was carried out on the resultant waveforms in order to examine the power spectrum of the EEG waveforms.

3.3 Results

3.3.1 Experiment 1: Behavioural results

Subjective reports obtained from participants when listening to a sequence with a large frequency separation (greater than 6 semitones) between the low and high tones indicated that the spontaneous percept for the majority of participants (10/15) was of the high tone in the right ear alternating with the low tone in the left ear (RHi/LLo). The remaining five participants reported hearing the low tone in the right ear, alternating with a high tone in the left ear (RLo/LHi).

For the 15- and 20-semitone frequency separations, the subjective reports after priming indicated that participants were able to change their perception of the sequence, depending on the priming sequence. For example, participants with the spontaneous perception of RLo/LHi reported hearing the reversed percept of RHi/LLo if the priming sequence was either high tones in the right ear or low tones in the left ear. In contrast, the subjective reports for the two smaller frequency separations (1 and 6 semitones) suggested that participants perceived a fused stream and that they were not able to precisely locate the ear in which they heard the low and high tones.

In the detection tasks, listeners' sensitivity to the deviant target was estimated by calculating d' for the detection of deviants for all conditions. The value of d' here and elsewhere was calculated by subtracting the inverse cumulative standard normal distribution function of the proportion of false alarms (participant responses to trials in which there was no deviant in the target stream, as a proportion of all trials with no deviant in the target stream) from the inverse standard normal cumulative distribution function of the proportion of hits (participant responses to trials in which there was a deviant in the target stream, as a proportion of all trials in which a deviant was present in the target stream): $d' = z(H) - z(F)$. To test whether performance was higher in the listening condition that listeners seem to report naturally, a one-way ANOVA on the overall d’ scores was performed with listening condition (RLo, RHi, LLo, LHi) as the factor. No significant difference was found between the four listening conditions $F(3,12)=3.28$, $p>0.05$.

Two predictions can be made if stream segregation plays a role in determining performance in this task. First, segregation is known to increase with increasing frequency separation (van Noorden, 1977; Miller and Heise, 1950); therefore, improved performance would be expected with increasing frequency separation between the two tones. Second, stream segregation tends to build up over time (Anstis and Saida, 1985; Bregman, 1978); therefore, performance should improve over the duration of each sequence, at least for frequency separations at which build-up is expected. The data are generally consistent with the first prediction, with overall performance increasing with increasing frequency separation from 1 semitone to 20 semitones (Figure 3.3 shows that the d’ scores increase for the higher frequency separations). The data are also consistent with the second prediction, with better performance observed during the latest than the earliest time periods, at least at the two larger frequency separations (Figure 3.3). These trends were confirmed by a repeated-measures ANOVA on the d’ values, with frequency separation (four levels) and temporal position of the target deviant (3 levels) as factors. The ANOVA confirmed a main effect of frequency separation, $F(3,12)=122.8$, $p<0.001$, as well as a significant linear trend $F(1,14)=222.8$, $p<0.001$, supporting the hypothesis of increasing performance with increasing frequency separation. The main effect of temporal position was also significant, $F(2,13)=20.13$, $p<0.001$), as was the interaction between frequency separation and temporal position $F(6,9)=10.04$, $p<0.005$. Post hoc tests indicated a significant difference in the d’ scores for early deviants compared to mid and late deviants for the 15 and 20 semitone conditions, $p<0.001$.

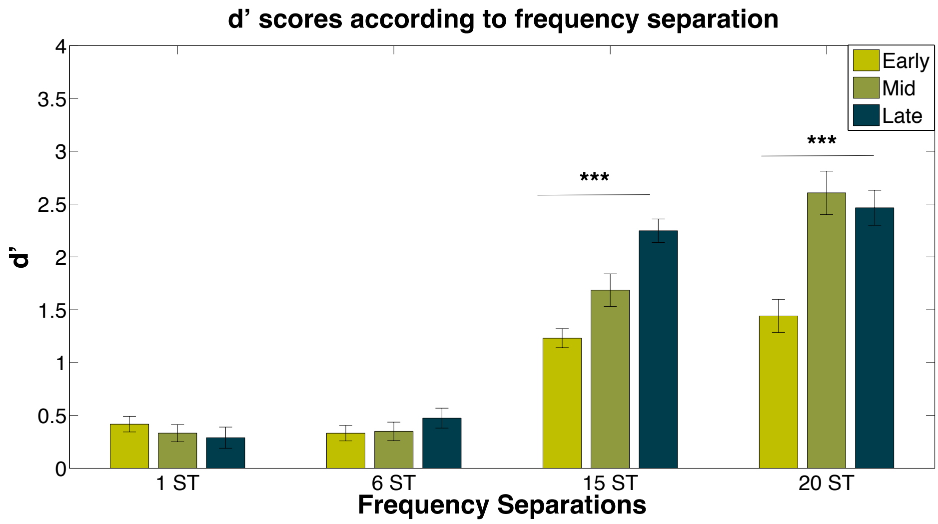


Figure 3.3: *Average deviant detection scores across four different frequency separations. The three differentially coloured bars per frequency separation indicate the detection scores (d') for the early, mid and late deviants. For the higher frequency separations, a significant increase in the detection scores of the late deviants compared to the early deviants was found.*

3.3.2 Experiment 2: Combined EEG and behavioural results

The behavioural results, averaged across the four conditions (RLo, RHi, LLo, LHi) for the single frequency separation (1000 and 3000 Hz, or about 19 semitones), are shown in Figure 3.4. Similar to the results obtained in Experiment 1, a significant difference was seen between the deviant detection d' scores for the early and late target positions, $F(1,9)=9.56$, $p<0.01$.

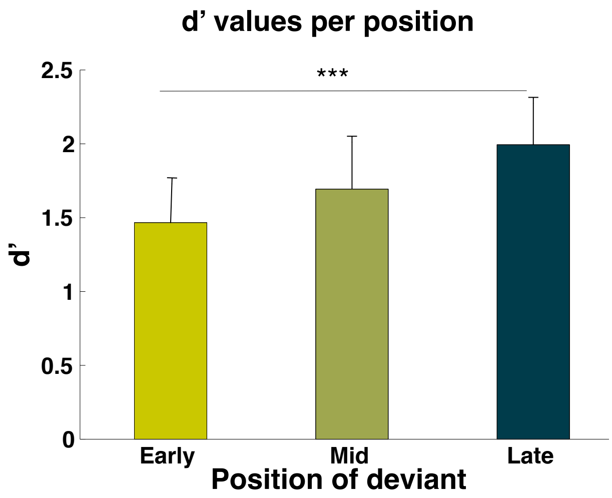


Figure 3.4: *Average deviant detection results from behavioural data obtained in Experiment 2 averaged over 10 participants. Data showed a significant difference between d' scores for early and late deviants.*

As described in the Methods, the EEG signals from two of the conditions (RLo and LHi) were averaged and subtracted from the sum of the EEG signals from the other two conditions (RHi and LLo) to enhance the difference between conditions in which participants attended to different time epochs. The prediction was that high activity at 3 Hz (the repetition rate of the target tones) would indicate enhancement of the attended tones. It was found that for the correct trials (all correct trials as well as the subset of correct trials taken to match the number of incorrect trials; see Figure 3.5 and Figure 3.6 – top panel), 3-Hz activation emerged prominently during stimulus presentation, whereas it was absent during the trials that were incorrectly responded to (Figure 3.6 – bottom panel).

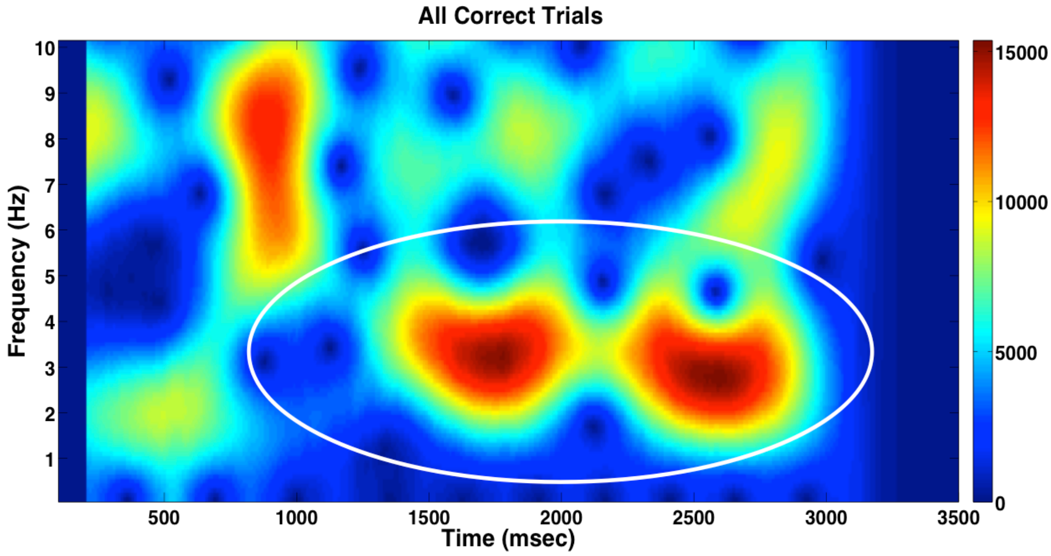


Figure 3.5: *Spectral analysis of all correct data indicating a 3 Hz pattern (data combined across all 10 participants and all conditions). The spectral analysis was carried out on the averaged waveform across the four priming conditions as described in the methods section (the averaged waveform of conditions RLo and LHi were subtracted from the averaged waveforms of RHi and LLo).*

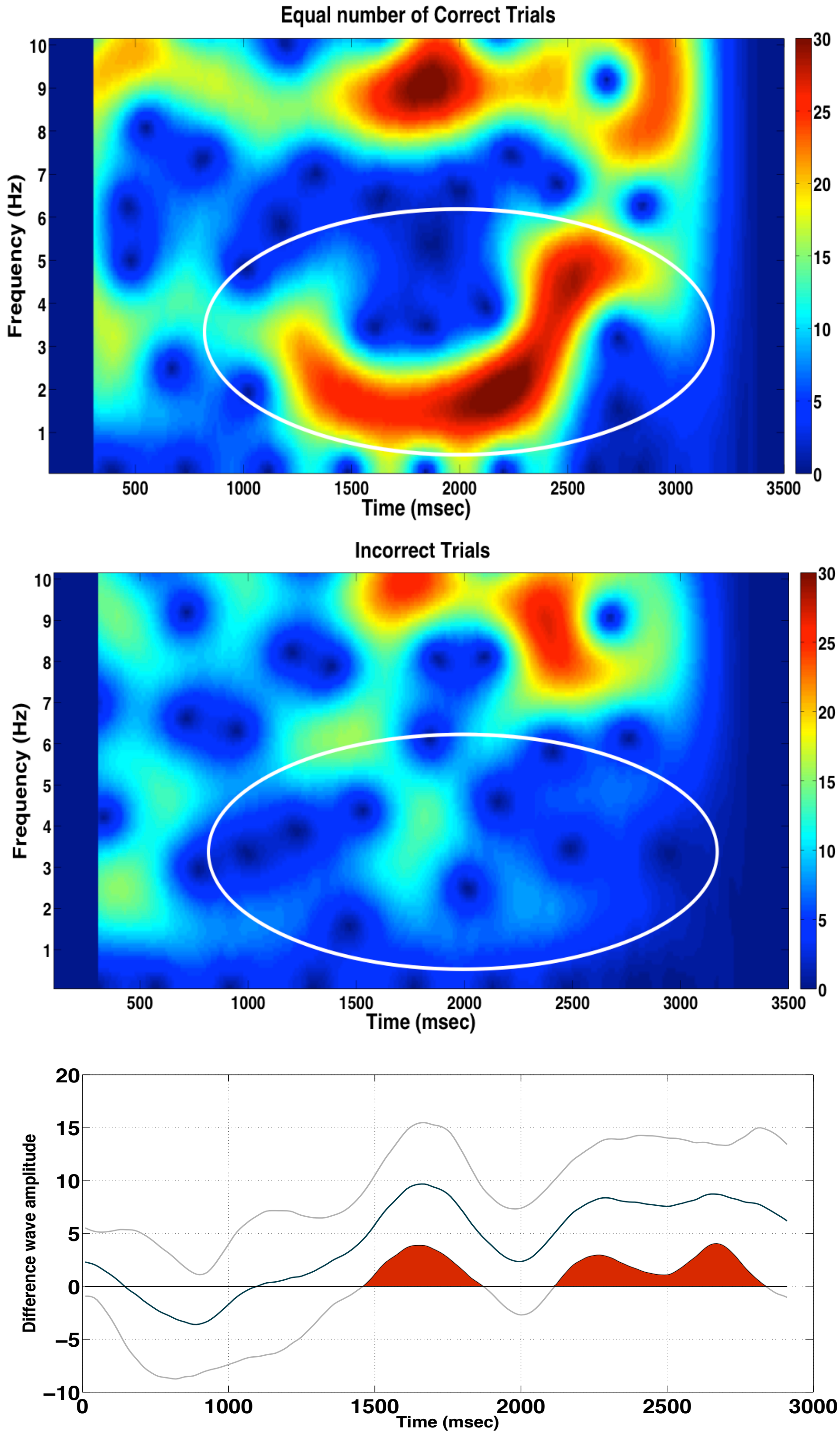


Figure 3.6: *Subtracted waveforms Spectral analysis of equal number of correct and incorrect trials indicating a 3 Hz pattern (data combined across all participants and all conditions) for the correct trials (top panel) but not for the incorrect trials (middle panel). Bottom panel shows the bootstrapped difference between the correct and incorrect trials. The parts highlighted in orange show the regions of significant difference (p<0.001).*

To confirm that the 3-Hz activation in the difference spectrograms was significant, a repeated-measures analysis was used in which, for each participant, the time-frequency series for each participant's differences were subjected to bootstrap analysis (1000 iterations, balanced; Efron & Tibshirani, 1994). At each time point, the proportion of iterations below or above the zero line was counted. If the proportion was less than 0.1% or more than 99.9% for 5 adjacent samples (50 ms), the difference was judged to be significant. As shown in Figure 3.6, the activity in the 3-Hz region for the correct and incorrect trials was significantly different.

Aside from the response to tones in the main sequences, the responses during and after the priming sequence were also examined. The priming sequence of three tones was presented at 6 Hz. The magnitude of 6-Hz activity from the onset of the priming sequence until the offset of the test sequence for all the incorrect trials as well as an equal number of correct trials was calculated. The 6-Hz activity during the silent period showed significantly greater suppression for the correct trials than the incorrect trials (Figure 3.7). Interestingly, the 6-Hz activation after the test sequence onset seemed higher for the correct trials compared to the incorrect trials, although that difference did not reach significance (Figure 3.7). In the later part of the sequence, the correct trials showed a decrease in the 6-Hz activation, in line with the increase in 3-Hz activation observed for these trials.

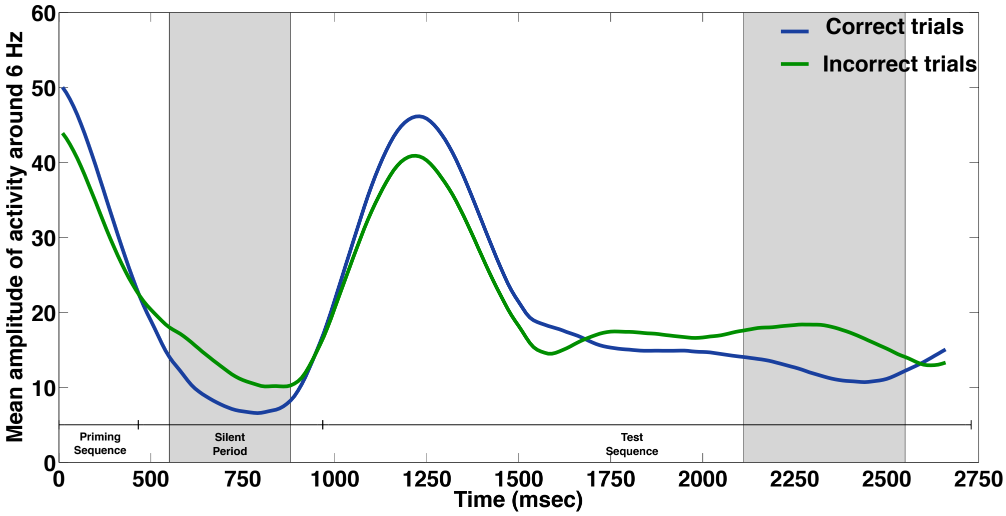


Figure 3.7: *Plot indicating the magnitude of 6-Hz activity (presentation frequency of the probe tone). Grey bars indicate regions of significance (p<0.005) showing suppression for correct trials in the silent period and towards the end of the sequence.*

3.4 Discussion

The two experiments described in this chapter investigated the percept elicited by a complex stimulus of alternating high and low tones played in opposite presentation phases in the two ears, known as Deutsch's octave illusion (Deutsch, 1974). The questions asked were whether the illusion could be understood in terms of the basic principles of auditory streaming, whether the perception of the illusion could be manipulated by directed attention by changes in listening instructions provided via auditory priming cues, and whether the corresponding neural activity mirrored these changes in perception of the illusion. The results provide support for all three hypotheses.

Role of stream segregation

The octave illusion is thought to arise from mechanisms involving concurrent and sequential sound segregation. As noted earlier, there are certain key distinct properties of stream segregation. The properties that the current two experiments tap into are the build-up of segregation over time as well as the effect of frequency separation on streaming. As previously noted by Brancucci et al. (2009), the fact that the illusion breaks down at small frequency differences suggests that it is mediated at least in part by auditory streaming constraints. As has been previously noted, stream segregation breaks down if the frequency separation is too small (van Noorden, 1975). The behavioural results from experiment 1 confirm and extend these observations by showing a deterioration in a performance-based task in conditions with a small frequency difference between the low and high tones (of 6 semitones or less), suggesting a lack of stream segregation that results in an inability to “hear out” and follow a subset of tones within the complex sequence.

Another key indicator of streaming is a build-up of segregation over time as the sequence unfolds (Anstis and Saida, 1985). A build-up effect was observed when the frequency separation between the tones was large enough for participants to perform well in the deviant detection task (15- and 20-semitone conditions). The build-up appears more rapid in the 20- than the 15-semitone condition, in line with earlier work showing a very rapid build-up at large separations (Micheyl et al., 2007). Thus the behavioural results are consistent with the hypothesis that the Deutsch illusion is subserved by the same mechanisms that govern auditory streaming. This is based on the two characteristics observed for the octave illusion that are in line with the parameters of stream segregation.

Effects of priming and directed attention on perception and EEG responses

Using stimuli very similar to the ones used in the experiments described in this chapter, Deutsch reported an auditory illusion, whereby an alternating sequence of low and high tones in both ears (Figure 3.1, Stimulus) were heard as a series of alternating tones, with low tones in one ear, and the high tones in the other ear, presented at a rate that was half that of the actual presentation rate (Figure 3.1, Percept). It appeared as if only the tones from one ear were being perceived, but that one of the frequencies was being mislocated to the opposite ear (Deutsch, 1974; Deutsch and Roll, 1976; Deutsch and Gregory, 1978). The behavioural results from the current experiments extend those original findings by demonstrating that instructing listeners, via a priming sequence, to attend to a particular tone frequency in a particular ear, could alter the subjective percept, as well as performance-based measures. For e.g. if a participant’s unbiased percept of the illusion is Right ear - Low tones and Left ear – High tones, the participant can also as easily perceive the inverse percept of Right ear - High tones and Left ear – Low tones if cued to either high tones in the right ear or low tones in the left ear. This showed that the illusion is robust but malleable to instructions and attention.

The simultaneously gathered data from EEG activity also indicated that participants were able to attend to the target tones in the correct ear, which were presented at half the rate of the stimulus i.e. 3 Hz. Thus, consistent with the reported perception, in trials where the participants were able to detect the deviant in the target stream, neural activity at the target repetition rate (3 Hz) was enhanced, in phase with the target presentation times. Interestingly, in trials where the participants were not successful in following the target tones (as evidenced by failure to detect the target deviant), no such 3-Hz activity was detected, leading to a significant difference in 3-Hz activity between incorrect and correct trials. This difference in 3-Hz activity was not due to the larger number of correct trials (leading to potential enhancement of the signal-to-noise ratio), because the difference was still observed when the same numbers of trials were evaluated in both correct and incorrect categories. This indicated that the EEG activity, even for a smaller, randomly chosen subset of the correct trials, robustly indicates the attended tone streams and resulting response.

This enhancement of EEG activity associated with the attended stream of tones is consistent with a growing body of literature showing enhanced responses to attending (and detected) streams in a background of other streams (Alain et al., 2001; Alain & Izenberg, 2003; Carlyon, 2004; Carlyon et al., 2001; Cusack et al., 2004; Dyson & Alain, 2004; Gutschalk et al., 2005, 2007; Hillyard et al., 1973; Zion Golumbic et al., 2013).

Lastly, the time-course of the EEG power in the 6-Hz band was examined. This rate corresponded to the rate of tone sequences both during the priming and subsequent test sequence. Linke et al. (2011) found that in a task where a priming sequence precedes a test sequence, a frequency specific suppression in the maintenance period (the period between the priming sequence and the test sequence) is seen. They suggested that this mechanism suppressed activity that could interfere with the representation of the priming sequence. The related secondary hypothesis for this experiment was that more effective priming would leave a stronger suppression of activity and that a stronger suppression in the priming sequence presentation frequency in the correct trials as opposed to the incorrect trials would be seen. This hypothesis was confirmed by the finding of stronger suppression in the 6 Hz frequency region for "correct-trials" compared to "incorrect trials".

The octave illusion as a probe of multistable perception and perceptual organization

Studies of the perception of, and neural responses to, multistable stimuli can help us understand how objects or sources in the environment with conflicting or ambiguous cues are grouped according to specific characteristics to form a coherent representation of our surroundings (Schwartz et al., 2012). Several theories regarding the principles underlying perceptual bistability and multistability have been put forward. Leopold and Logothetis (1999) have suggested that a 'central, supramodal mechanism' underlies the perceptual decision making in multistable stimuli. Tong et al. (2006) proposed another model using multistable stimuli in the visual domain with a focus on the idea of distributed competition and have suggested that it is essential to understand the underlying neural mechanism involved in the processing of multistable stimuli, perceptual grouping and the effect of attention on them.

The multistable stimulus used initially by Deutsch (1974) has been studied in different contexts and the robustness of the percept has been investigated across a variety of parameters. It has already been demonstrated that the percept of this illusion is robust to changes in tone duration (Zwicker, 1984), intensity (Deutsch, 1978), frequency separation (Brancucci et al., 2009), and timbre (McClurkin and Hall, 1981), and can also be elicited by aperiodic stimuli like band-pass noise (Brännström & Nilsson, 2011). It was noted by Deutsch and Roll (1976), and later confirmed by Brancucci et al. (2009), that the illusion is not dependent on the tones being in exact octave relationship. In fact, Brancucci et al. (2009)

reported that the illusory percept was dominant for all musical intervals tested that were larger than a perfect fourth (roughly a ratio of 4:3 or a frequency difference of 33%).

In neuroimaging studies using this stimulus, either listeners' spontaneous percepts were tested beforehand and recordings were passively recorded without instructions on what to attend to within the sound sequences (Lamminmäki and Hari, 2000; Brancucci et al., 2012; Lamminmäki et al., 2012) or the response measures for a task-based study were mainly focused on the listeners' subjective responses regarding their percept (Brancucci et al., 2014). In contrast, our study investigated the effects of actively guiding the perception of the stimulus between different possible perceptual organizations. By including an active, objective, deviant detection task, a greater level of control could be exerted over the listeners' percept for every trial.

Some hypotheses regarding the nature of this illusion have been explored. Deutsch and Roll (1976) suggested that listeners generally reported the tone frequencies that were presented to their "dominant" ear (usually the right), through suppression of the non-dominant ear. Such suppression was postulated not to occur for sound localisation, but instead the higher tone tended to be localised to the right, and the lower tone localised to the left, regardless of the ear of presentation. The present set of experiments also found that the illusion is robust to stimulus variations (with the except of very small frequency separations, where streaming breaks down), but also that the spontaneously reported perceptual organization was very malleable to instructions, as provided by the priming tones. In fact, with the priming tones present, no advantage was observed for the spontaneously perceived organization over the other alternatives. If the stimulus induced an illusory percept that could not be overridden by directed attention, the participants would have only been able to carry out the deviant detection task efficiently in two of the four attention conditions; instead, performance was equivalent across all four conditions. Indeed, the malleability of the perception of this ambiguous stimulus renders it as a highly promising tool with which to study further the perception and neural correlates of auditory stream segregation for a stimulus that involves both sequential as well as synchronous sound segregation.

Chapters 4 and 5 aim to further explore this stimulus paradigm as it is vital to further explore the mechanisms as well as characteristics of this illusory paradigm from a streaming perspective. This illusion has been extensively studied from the point of view of the current dual-mechanism theory. However, considering the current findings of some of its mechanisms (if not all) having characteristics similar to a complex stream segregation paradigm, it is imperative to try and understand the nuances of this stimulus.

Chapter 4

Summary

This chapter continues to explore the dynamics of the octave illusion stimulus introduced in Chapter 3. The results of the studies in chapter 3 indicated that the stimulus sequence, similar to the octave illusion, was indeed governed by the principles of stream segregation. Additionally, the role of directed attention was established by indicating that the percept of the illusion can be manipulated by directed attention. The dual mechanism model suggested by Deutsch (1981) to explain the octave illusion suggests that the alternating tones heard in the octave illusion were the tones present in the dominant ear while the tones in the opposite ear were suppressed. In this chapter, two psychophysical paradigms were used to test this explanation of the mechanism of the octave illusion. The alternative hypothesis tested was that the percept of the illusion is determined by the concurrent tones in the stimulus. For example, if the illusory percept is Right ear - Low tones and Left ear - High tones, the alternative hypothesis suggests that low tones in the right ear and the high tones in the left ear are the ones heard in the illusory percept as opposed to all the high and low tones of any particular dominant ear (as would be the case in the dual-mechanism model). The results of the psychophysical

studies indicate that the concurrent tones seem to contribute to the pitch of the tones in the octave illusion percept.

4.1 Introduction

In chapter 3, a new perceptually multistable stimulus, based on a variant of the Deutsch's 'octave illusion' (Deutsch. 1974) was investigated to study stream segregation (see Figure 3.2 – Chapter 3). The results of Chapter 3 suggested that stream segregation plays a role in the perception of this illusion. Furthermore, it was established that the illusion could be manipulated by directing attention towards one of the two possible percepts (either perceiving high tones in the right ear and low tones in the left ear or vice versa). The first experiment in chapter 3 investigated whether the illusion involved a role of stream segregation. If so, it was hypothesised that the illusory percept depended on segregation of sources and hence would be easier to parse apart with an increase in frequency separation between the tones. The results revealed a statistically significant increase in stream segregation with increase in frequency separation between the low and high frequency tones of the octave illusion stimulus. Furthermore, a significant effect of build-up was measured using an objective, deviant detection task (detecting an amplitude increase in the target stream), which further indicated the role of stream segregation. The second experiment in chapter 3 used EEG recordings to indicate that the percept could be modified effectively by directing attention using a precursor tone sequence. The results of the two experiments described in Chapter 3 indicated that the octave illusion could be a powerful paradigm to study a complex, stream segregation scenario where both sequential as well as concurrent sound segregation are potentially involved. However, as the illusion had not been studied in detail from a streaming perspective, it is crucial to re-establish and understand the mechanisms of the illusion further. We start by understanding the dual-mechanism model suggested by Deutsch (1981) in detail.

4.2 Deutsch's dual mechanism model

The stimulus used to elicit the octave illusion is shown in the upper portion of Figure 4.1. Two pure tones that were spaced an octave apart were repeatedly presented in alternation. The same two-tone sequence was played simultaneously to both ears but one tone out of step with each other. This means that when the right ear receives the high tone, the left ear receives the low tone, and vice versa.

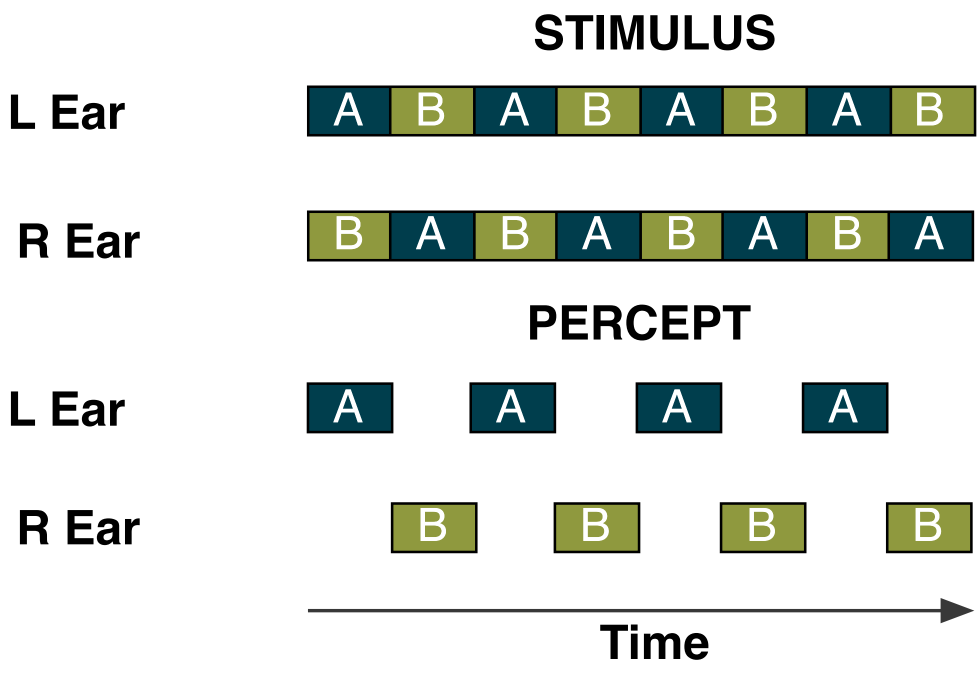


Figure 4.1: *Schematic representation of the stimulus and resulting illusory percept in Deutsch's Octave Illusion (Deutsch, 1974). Tones A and B indicate pure tones of frequencies 400 and 800 Hz.*

The octave illusion can be perceived in a number of different forms (Deutsch, 1974, 1980, 1981, 1983, 1987, 1995). The majority of listeners hear a single tone that switches from ear to ear, while its pitch simultaneously shifts back and forth between high and low. For example, the typical illusory percept could be described as one ear perceiving the pattern “high tone - silence - high tone - silence” (in right-handers, this is generally the right ear) while the other is receiving the pattern “silence - low tone - silence - low tone” (in right-handers, this is generally the left ear). This percept is illustrated in the lower portion of Figure 4.1. This percept can be sustained even when the earphone positions are reversed; i.e. the apparent locations of the high and low tones often remain fixed irrespective of the physical stimulus being reversed. The tone (high or low) that was heard at half the rate of the stimulus in the right ear continues to appear in the right ear, and the tone that had appeared in the left ear continues to appear in the left ear.

Deutsch (1975) suggested that the octave illusion resulted from a combination of two separate decision mechanisms, one to determine the perceived pitch, and the other to determine the perceived location of the tone. The model is depicted in Figure 4.2.

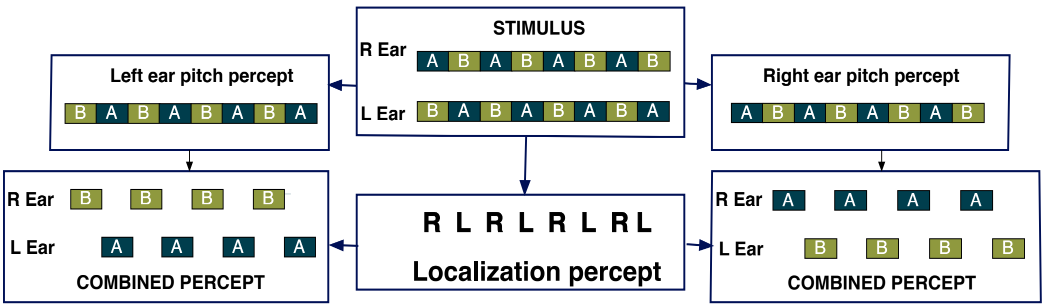


Figure 4.2: *Schematic diagram illustrating Deutsch’s dual-mechanism model consisting of two decision mechanisms, one to determine the perceived pitch and the other to determine the perceived location, both of which combine to produce the resultant illusory percept. To determine the perceived pitch, listeners attend to the pitch information in one ear and suppress the corresponding pitch information in the other ear. The perceived location of the tone based on the ear that received the higher frequency signal at that particular point in time. In the figure above, A = high frequency tone, B = low frequency tone. Adapted from Deutsch (1981).*

To provide the perceived pitches of the tones (i.e., the melodic line), the frequencies arriving at only one ear are followed, and all the frequencies arriving at the other ear are suppressed. However, each tone is localised at the ear that receives the higher frequency, regardless of whether a pitch corresponding to the higher or the lower frequency is in fact perceived.

To understand Deutsch's model further, consider an example case of a listener who follows the pitches delivered to his right ear. In the physical stimulus, when the high tone is presented to the right and the low tone to the left, this listener hears a high tone as the tones of only one ear (in this example, the right ear) are supposed to be heard according to the pitch mechanism of the Deutsch dual-mechanism model. The listener also localises the tone in the right ear, because this ear is receiving the higher frequency. However, when the low tone is presented to the right ear and the high tone to the left, this listener now hears a low tone, because this is presented to the right ear, but localises the tone to the left ear instead, because the left ear received the higher frequency. So the entire pattern is heard as a high tone to the right that alternates with a low tone to the left. It can be seen that, on this model, reversing the positions of the earphones would not alter the basic percept. However, for the case of a listener who follows the pitches presented to the left ear instead, holding the localisation rule constant, the identical pattern would be heard as a high tone to the left alternating with a low tone to the right. Follow-up experiments by Deutsch and colleagues have found evidence to support this model (Deutsch, 1978, 1980, 1981, 1987, 1988; Deutsch & Roll, 1976). It has also been established that the illusion can be perceived for acoustic parameters that are different from the initial parameters used by Deutsch (1974) to elicit the illusion. For example, it is known that the illusion can be elicited by narrow-band noise, complex tones, frequency modulated tones as well as tone sequences that are not separated by exactly an octave (Brancucci et al., 2009; Brännström & Nilsson, 2011; Deutsch & Roll, 1976; Lamminmäki et al., 2012).

Through the dual mechanism model suggested by Deutsch (1981), the octave illusion can be seen as a simple yet striking example of illusory feature conjunction; an instance when the spectral and spatial features relating to an incoming stimulus obtained by the perceptual system are faultily bound together to give rise to an illusory percept. This incorrect feature conjunction throws open the crucial issue of how humans usually manage to arrive at correct binding solutions. Deutsch (1980, 1981) provided an early dual mechanism parallel processing model of how correct binding can occur for the case of values of two attributes (pitch and location) and that also explains the incorrect binding that occurs in the octave illusion (see Figure 4.2). This model has been expanded to account for the correct binding of values of three or more attributes, such as pitch, location, and loudness (Deutsch, 1988) especially in the case of the scale illusion (Deutsch, 1975).

Focusing on the octave illusion in particular, grouping in this illusion according to the Deutsch dual mechanism model is based on spatial location. This indicates that the pitches that are heard correspond to the tones presented either to the listener's right ear or to his left.

Deutsch and colleagues have suggested that the reason for such a perceptual strategy to evolve is that it enables the listener to follow new, on-going information with the minimum of interference from echoes or reverberation (Deutsch, 1998). In everyday listening, when the same frequency emanates successively from two different regions of space, the second occurrence may well be due to an echo. So a useful perceptual strategy involves suppression of the second occurrence of the sound from conscious perception. A similar argument has been advanced for the precedence effect: in listening to music, a single sound image may be obtained when the waveforms arriving from two different spatial locations are separated by time intervals of less than around 70 ms (see also Wallach et al., 1949; Zurek, 1987).

4.3 Need for exploring the dual-mechanism model

The octave illusion has been hard to explain and over the past few years; it has been noted that the illusion does not strictly adhere to an octave relationship between the tones (Brancucci et al., 2009; Brännström & Nilsson, 2011). Previous attempts at explaining the illusion strictly based on the frequency relationships of the tones hence do not hold ground (Bregman and Steiger, 1980; Chambers et al., 2002). The dual mechanism model (Deutsch, 1981) has been questioned (Chambers et al., 2002, 2004) as well as supported (Lamminmäki et al., 2010) but does not explain fully some of the points raised in this section.

Bregman & Steiger (1980) had suggested that in the case of the classic octave illusion (see Figure 4.1), the auditory system treated the 800-Hz tone as a harmonic of the 400-Hz tone and thus localised the percept of the tone at the ear receiving the "more reliable higher harmonic". Chambers et al. (2002) suggested that the octave illusion percept was based on dichotic fusion, which meant that the percept was made of the tones from both the ears fusing to form a percept that varied very slightly in overall perceived frequency. Although the Chambers et al. (2002) study highlighted the aspect of bilateral grouping of tones, it has been established that the tones perceived do not sound like a fused auditory image and in fact, correspond to the pitch of the component high and low frequency tones (Deutsch, 2004). In particular, the theory of dichotic fusion would also have to be rejected, as it would predict that the illusion would persist even for tones having similar frequencies in the dichotic pair. However, the results by Brancucci et al. (2009) as well as the psychophysical results of Chapter 3 – Experiment 1 demonstrate that with similar frequencies (e.g., minor third interval or with frequency separations of 1 or 6 semitones), the illusory percept almost disappears.

From a streaming perspective, the dual mechanism model does not agree with the streaming model of temporal coherence (Elhilali et al., 2009; Shamma et al., 2011). For example, the low and high tones in the stimulus are temporally coherent (see Figure 4.1) and in theory, segregate for a frequency separation of 12 semitones or lesser is difficult to obtain and it is even more unlikely that one of the components of these temporally coherent units is completely suppressed (for example, suppressing Right Low in the coherent tone pair of Right Low and Left High).

Furthermore, the dual mechanism model indicates that the pitch percept of the illusion corresponds to the frequency sequence present in the 'dominant' ear of the individual. If this were the case, then directing attention to the non-dominant ear of the listeners should not change the percept of the listeners. However, it was found that the percept for all listeners can be manipulated by a simple precursor sequence (as indicated by the results of Chapter 3).

Since the percept is malleable, the concept of localising to the ear receiving the high tone as well as the dominant ear theory does not adequately explain the percept associated with the illusion.

It thus was important to understand the role of the coherent tones (present in the presumed suppressed non-dominant ear) and the acoustical parameter limits within which the illusion breaks down or is maintained.

As seen in Figure 4.3, the illusion paradigm can be thought of as two coherent-tone pairs (X and Y) alternating throughout the stimulus. The main questions that come to mind are:

1) What happens to the illusory percept if Y is removed?

2) Do the acoustical properties of Y matter?

3) How does the illusion get affected if the tones in the suppressed ear (which according to the Deutsch model are not heard) are changed?

Several pilot experiments were conducted (described in section 4.4) to understand the questions listed above. It has already been established that the illusion occurs for a larger frequency separation in chapter 3 and an exact octave separation is not needed (Brancucci et al., 2009). All experiments were therefore done keeping the frequency separation between the tones as 19 semitones.

4.4 Pilot observations

The following sets of observations were made from pilot behavioural experiments carried out to explore various properties of this illusion. Four listeners were tested on a single block of thirty trials. The percept for all four pilot paradigms for all the four listeners was identical. The general percept for each condition has been reported below. All the low tones (marked ‘A’ in Figure 4.3) were pure tones of 1000 Hz and all high tones (marked ‘B’ in Figure 4.2) were pure tones of 2996 Hz (frequency difference of 19 semitones). Each tone was 100 ms long (including 10 ms onset and offset cosine ramps) and the inter-tone intervals were 50 ms long. The sequences were generated in MATLAB (MathWorks Inc. Natick, MA, USA) and were presented at a sampling rate of 44.1 kHz. For the following pilot observations, consider the typical stimulus as being made of alternating, coherent-tone *pairs* (Figure 4.2). For all further descriptions, the coherent tone pairs shall be referred to as X and Y.

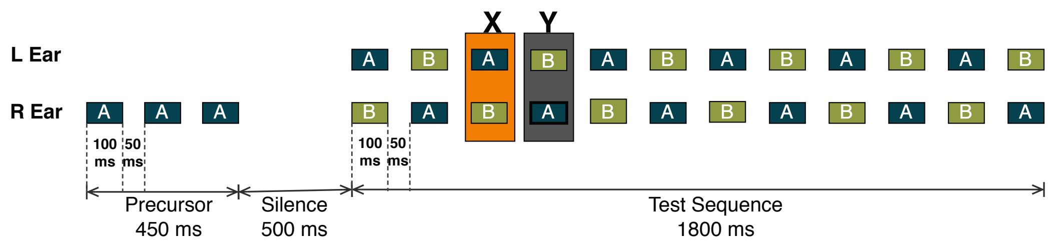


Figure 4.3: *Schematic representation to illustrate the coherent pairs described in section 4.2. The diagram indicates that X and Y represent a coherent tone pair and that this paradigm can be also understood as alternating, opposing coherent tone pairs. A and B are low and high frequency pure tones respectively. Each tone was 100 ms in length with a 50 ms silent gap between tones. A precursor tone sequence consisting of three tones was presented before the test sequence to indicate the frequency and ear the listener needs to attend to in a particular trial (in the sample trial shown, the precursor is a low frequency tone sequence in the right ear).*

Observation 1: The first, most basic observation was that the illusory percept could not be elicited with only X or only Y repeated over time. For example if only X was repeated over time, it would be a typical stimulus used in concurrent sound segregation with all the A tones in the left ear and all the B tones in the right ear occurring simultaneously. The percept for all the listeners in this case was a fused complex tone percept.

Observation 2: The illusion also ceased to exist when a noise band was used in place of either Y (or X). The percept of only X alternated with a white noise burst was that of a fused complex tone alternating with a noise burst. This paradigm was tested to rule out whether the illusion could be elicited simply by putting any acoustical energy in the alternating paradigm (i.e. with Y being white noise) or it required a more specific configuration of Y.

Observations 1 and 2 indicate that the illusion not only requires a sound alternating with 'X', it also requires a specific configuration of Y to occur in order to sound like the octave illusion. Observation 3 aims to understand the properties of 'Y' that lead to the percept of the octave illusion.

Observation 3: Following on from observation 2, a pilot experiment was carried out when the stimulus was changed dynamically. X was kept constant whereas Y started off being the same as X. Gradually; the frequencies of the component tones of Y were changed so as to approximate the inverse pattern of X (i.e. like the typical Y tone pair that is the inverse of X). For example, in the stimulus used, A =1000 Hz, B=2996 Hz. The stimulus started off with a constant sequence of X and Y where X and Y were the same. As the stimulus progressed, the Y tone pair in the left ear was gradually increased in frequency from 1000 Hz to 2996 Hz (i.e. turning A into B) and the Y tone pair in the right ear was gradually decreased in frequency from 2996 Hz to 1000 Hz (i.e. turning from B to A) in gradual steps of 75 Hz. The aim was to examine if the illusion required only a slight change between X and Y or it required X and Y to be exactly opposite. The observation made was that for the initial parts of this sequence, the illusion was not heard. However, only towards the end of the latter half of the sequence, when the frequency configuration of Y started approximating the inverse of X (within 150 Hz of the target Y configuration), the illusion was perceived. However, the illusion was maximally perceived when Y was exactly the inverse configuration of X. This paradigm was extremely interesting in terms of understanding underlying context effects in the perception of the illusion. In simpler terms, it aimed to investigate what the configuration of X and Y needed to be in order to not be perceived as fused complex tones but sound like segregated illusory pure tones.

Building on this further, it was crucial to understand the role of the tones presented in the 'suppressed' ear. This was to check if the pitch percept was a result of the frequencies presented sequentially (i.e. all the tones in one ear – the dominant ear) as suggested by the Deutsch model (1983) or not. Observation 4 was a pilot experiment that addressed this aspect.

Observation 4: This was an observation that led to the two psychophysical paradigms carried out below. In pilot 4, one of the tones from the standard octave illusion was removed either from X or from Y. For example, in Figure 4.3, the high tones (B) either from tone-pair X or tone-pair Y could be removed while the sequence was on going. In this particular example of the trial depicted in Figure 4.3, the listener's default percept at the start of the sequence would be right ear- low tones and left ear – high tones. According to the dual-mechanism model of Deutsch (1981), it was expected that the pitch was determined by the tones in the same ear (in this case for example, the right ear), and if the tone removed belonged to the sequence in the right ear (dominant ear), the percept would be disrupted. In contrast, any changes to the contralateral tones (for example, high tones in the Left ear) would not affect the pitch percept of the illusion. However, contrary to the prediction based on the dual mechanism model, the pilot results unanimously suggested that the deletion of the tones in the opposite ear resulted in disruption of the percept whereas the deletion of the tones in the same ear did not affect the percept of the illusion much.

Observation 4 led to the following two paradigms that were designed in order to answer the question: Which low and high tones contribute to the pitch percept in the octave illusion? For example, consider the trial shown in Figure 4.3. The hypothesis according to the dual mechanism model would indicate that the pitch percept arises from the right ear low tones and the right ear high tones. The alternative hypothesis was that the pitch percept arises from the right ear low tones and the left ear high tones.

4.5 Methods

4.5.1 Participants

Fifteen adults (21–30 years of age, 6 male) with normal hearing (audiometric thresholds <20dBHL from 250 to 8000 Hz) and reportedly no history of neurological disorders participated in the experiment. Informed consent was obtained after the procedures of the experiment were explained to them. The University College London Ethics Committee approved the experiment. Participants received payment for their participation. All listeners were naïve to the stimulus (none of them took part in the pilot experiments or any previous experiments) and were chosen on the basis of their ability to perceive the octave illusion without any precursors. All 15 listeners were also able to manipulate their percept of the illusion according to the four different precursor tones. For example, if the listeners were exposed to a trial with a precursor of low tones in the left ear, their percept would be low tones in the left ear and high tones in the right ear. Furthermore, the same listener would be able to change their percept to perceiving the low tones in the right ear and high tones in the left ear if the precursor of low tones was presented in the right ear.

4.5.2 Stimulus and Paradigms

Aim of Paradigms 1 and 2

The aim of both the stimulus configurations in this experiment was to determine which low and high tones of the stimulus were contributing to the low and high tone percept of the octave illusion.

Paradigm 1: Amplitude modulated tones

This paradigm consisted of a stimulus similar to the one used in Chapter 3 consisting of three Low or High precursor tones either in the left or right ear to indicate what side and frequency listeners needed to attend to. For each trial, one of the non-cued streams (i.e. not the same frequency as the precursor) was amplitude modulated. For e.g. Figure 4.4 shows an example trial of Paradigm 1. The listener is cued to the low tones in the right ear via a precursor sequence made of three unmodulated low frequency tones (for all trials, the precursor tones were always unmodulated). The left and right ears were presented with sequences of alternating low and high frequency tones. The illusory percept of the listener is low tones in the right ear and high tones in the left ear. The left ear-high tones were amplitude modulated (modulation depth of 0.75) whereas all other tones were unmodulated. The listeners' task was to report via a two-alternative forced choice task using a computer keyboard button press whether the most salient tone alternating with the cued tone was modulated or not. In the schematic presented in Figure 4.4, if the listener perceived the illusion with amplitude modulated tones, it would mean that the illusory percept arose from the coherent tones. If the listener reported hearing pure tones, it would mean that the percept was determined from the non-coherent tones. The precursors and modulated tones were completely randomized. For each trial, either the coherent or non-coherent tone streams could be amplitude modulated. This resulted in a total of 8 conditions (four precursors – Right or Left ear/ Low or high frequency with two conditions - coherent tone stream modulated/non-coherent tone stream modulated each). The stimulus was randomized hence each type of non-target stream (coherent/non-coherent) was modulated for 50% of the trials.

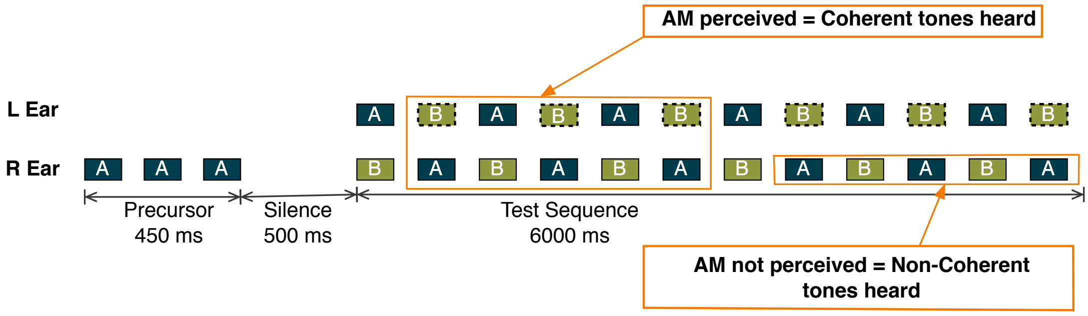


***Figure 4.4**: Schematic diagram illustrating paradigm 1. The black dotted lines indicate amplitude modulation (modulation depth of 0.75). A and B are pure tones with a 19 semitone separation (100 ms in length with a 50 ms silent gap between tones). In the example trial illustrated in the figure, the high tones (B) in the left ear are amplitude modulated. The precursor is a low frequency tone sequence in the right ear. For this trial, the listener has a percept of Right ear Low tones – Left ear high tones and the listeners are supposed to report whether the most salient high frequency tone alternating with the cued low frequency tone (Right Low) is amplitude modulated or not. In this particular example, if the listeners report hearing a modulated tone, it indicates that the coherent tone was heard and vice versa.*

Listeners also underwent thirty trials of distinguishing amplitude-modulated tones from un-modulated pure tones (carrier frequencies of the tones being 1000 Hz and 2996 Hz). In these trials, three tones of either amplitude modulated or pure tones were presented. The listeners had to indicate whether the stimulus was amplitude modulated or unmodulated pure via button press in a two alternative forced choice task. All listeners were able to achieve 100% accuracy on correctly identifying the amplitude-modulated tones.

All the low tones (marked 'A' in Figure 4.4) were pure tones of 1000 Hz and all high tones (marked 'B' in Figure 4.4) were pure tones of 2996 Hz (frequency difference of 19 semitones). Each tone was 100 ms long and the inter-tone intervals were 50 ms long. All tones had a 10 ms onset and offset cosine ramp. The modulation depth for the amplitude-modulated tones was set at 75 per cent. The sequences were generated in MATLAB (MathWorks Inc. Natick, MA, USA) and were presented at a sampling rate of 44.1 kHz. The experiment was presented using the Psychophysics Toolbox extension in MATLAB (Brainard, 1997; Pelli, 1997).

Paradigm 2: Fading tones

This paradigm was similar to paradigm 1 (Amplitude modulated tones) consisting of three Low or High precursor tones either in the left or right ear to indicate what side and sound frequency listeners needed to attend to. However, in this paradigm, for each trial, one of the non-cued streams (i.e. not the same frequency as the precursor) had a few tones in the middle of the stream reduced in amplitude. For e.g. Figure 4.5 shows an example trial of Paradigm 2. The listener is cued to the low tones in the right ear via a precursor sequence made of three low frequency pure tones. The left and right ears were presented with sequences of alternating low and high frequency tones. The illusory percept of the listener is low tones in the right ear and high tones in the left ear. The left ear-high tones were faded out and back in by halving the amplitude of the tones for each successive tone until the amplitude were 1/8th of the original amplitude after which the amplitude was increased in the same proportion. The change in amplitude occurred over a time-course of seven tones. The listeners' task was to report via a two-alternative forced choice task using a computer keyboard button press whether the most salient tone alternating with the cued tone changed in amplitude or not. In the schematic presented in Figure 4.5, if the listener perceived the illusion with a disruption in amplitude due to the fading out in one of the alternating tones, it would mean that the percept arose from the coherent tones. If the listener reported hearing uninterrupted tones without any change in tone amplitudes, it would mean that the percept was determined from the non-coherent tones. The precursors and faded tones were completely randomized and in each trial, either the coherent or non-coherent tone streams could be faded. The precursors and faded tones were completely randomized. For each trial, either the coherent or non-coherent tone streams could have the faded tones. This resulted in a total of 8 conditions (four precursors – Right or Left ear/ Low or high frequency with two conditions - coherent tone stream faded/non-coherent tone stream faded each). The stimulus was randomized hence each type of non-target stream (coherent/non-coherent) had the faded tones for 50% of the trials.

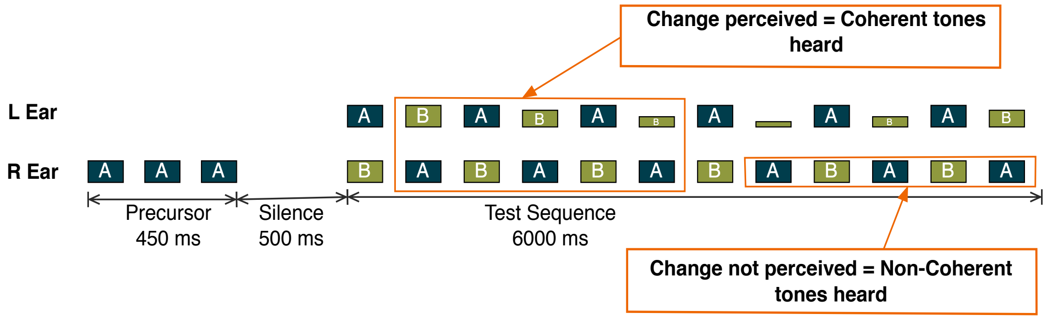


Figure 4.5: *Schematic diagram illustrating paradigm 2. The smaller tones indicate reduction of amplitude. A and B are pure tones at a 19 semitone separation (100 ms in length with a 50 ms silent gap between tones). In the example trial illustrated in the figure, the high tones (B) in the left ear are faded out and reintroduced. The precursor is a low frequency tone sequence in the right ear. For this trial, the listener has a percept of Right ear Low tones – Left ear high tones and the listeners are supposed to report whether the most salient high frequency tone alternating with the cued low frequency tone (Right Low) was faded out or not. In this particular example, if the listeners report hearing the tone fade away, it indicates that the coherent tone was heard and vice versa.*

All 15 listeners carried out both these paradigms. The total experiment with instructions, practice tasks and both paradigms took 2 hours. Each paradigm had 5 blocks with 12 test trials (60 trials per paradigm in total). Listeners were not given any feedback as that would result in a biased response.

4.6 Results

For each trial, the responses were recorded via button press according to whether the response was related to the state of the coherent or non-coherent tones. For e.g. in Figure 4.5, if the listener responded as ‘Change heard’, the response would be marked as a coherent-tone heard and vice versa. The trials were divided according to each precursor type as well to check if any listener was unable to do it for any particular precursor. Upon inspection, this was not seen for any listener. Next, the results were collapsed over all trials, as there was no difference noted for the individual precursors. Lastly, the results were scaled to range between -1 and +1. Any responses between -1 and 0 were meant to indicate that the responses consistent with the non-coherent tones were chosen. Any responses between +1 and 0 indicated that the results consistent with the coherent tones were reported. Results for 15 listeners are shown in Figure 4.6. The scatterplot in Figure 4.6 shows the results for all listeners on both the paradigms. Each dot indicates the scaled scores for a particular listener. Blue and green dots indicate performance scores for the two different paradigms. The Y-axis is scaled such that the top panel of the graph indicates the scaled scores denoting ‘coherent tones heard’ whereas the bottom panel of the graph indicates the scaled scores denoting ‘incoherent tones heard’. It can be observed that most listeners report listening to the percept corresponding to the coherent tones (data points would lie in the upper half of the plot). A one-sample t-test was conducted to test whether this trend of perceiving the coherent tone percept was statistically significant. The results showed a significant positive deviation from zero; $t(14) = 4.3566$, $p<0.0001$.

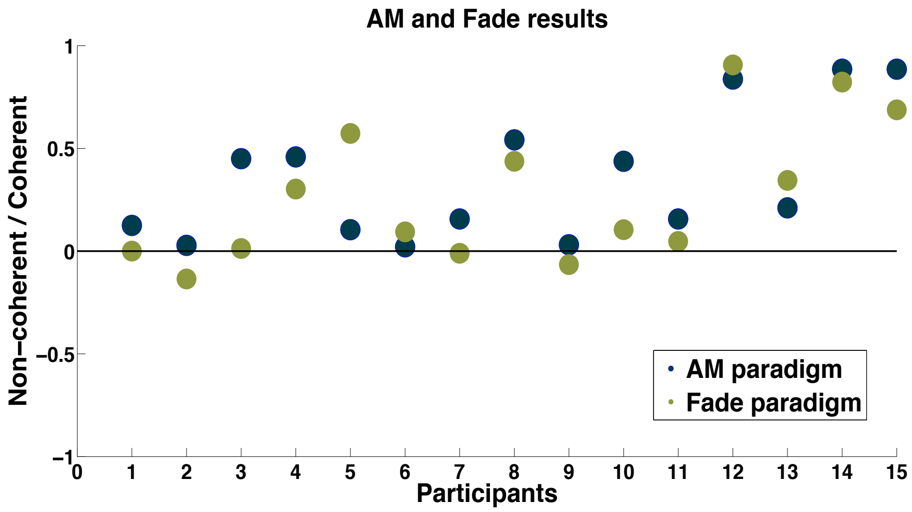


Figure 4.6: *Graph indicating results for 15 listeners for both paradigms. The dark blue circles indicate results for the amplitude modulated tone paradigm whereas the green circles indicate the results for the fading tones paradigm. The Y-axis is scaled such that the upper half of the graph (from 0 to +1) indicates when the response was 'coherent-tones heard' and the lower half of the graph (from 0 to -1) indicates when the response was 'incoherent-tones heard'. As seen from the statistics and the figure, for both paradigms (indicated by blue and green circles), most listeners responded to the coherent tone properties.*

4.7 Discussion

Both paradigms described above (amplitude modulated and fading tones) were designed to answer the question: Which of the low and high frequency tones contribute to the percept of the illusion? By the dual mechanism model proposed by Deutsch (illustrated in Figure 4.2), the pitch percept arises from the tones in either ear (deemed as the dominant ear) and that the tones are being misattributed in location. However, the results from paradigm 1 and 2 indicate that the percept of the illusion mostly arises from the concurrent tones. This would indicate that the illusory percept might arise from misattributions of the stimuli in *time* rather than location. The findings of this experiment, although unable to provide a complete explanation for the mechanisms of the illusion, suggest a previously unexplored aspect of this stimulus. The dual-mechanism model suggests that the incoherent tones (i.e. the tones of only one of the two ears) contribute to the pitch percept of the illusion. However, the data described above does not suggest the same. The dual-mechanism model does not provide for the coherent tones contributing to the percept and hence there is a need to try and alternately explain this illusion taking the findings of paradigms 1 and 2 into account.

Deutsch and Roll (1976) had carried out a set of experiments to understand which of the low and high frequency tones contributed to the pitch in the illusion. They used only right-handed listeners in the experiment. The listeners were presented with a stimulus sequence as shown in Figure 4.7 where a dichotic tonal sequence was presented in each trial. The basic configuration of the sequence consisted of three high frequency tones (800 Hz – indicated by the green ‘B’ squares) followed by two low frequency tones (400 Hz – indicated by the blue ‘A’ squares) to one ear. Simultaneously, the opposite pattern (three 400 Hz tones followed by two 800 Hz tones) was presented in the other ear. All the tones were 250 ms long and were separated by 250-ms silences. Their results suggested that for any stimulus presentation, most participants reported the pitch of the sequence delivered to one of the ears and the pitch of the other ear was ignored or suppressed. This led them to suggest that the tone sequence heard resulted from one of the ears rather than from the coherent tones.

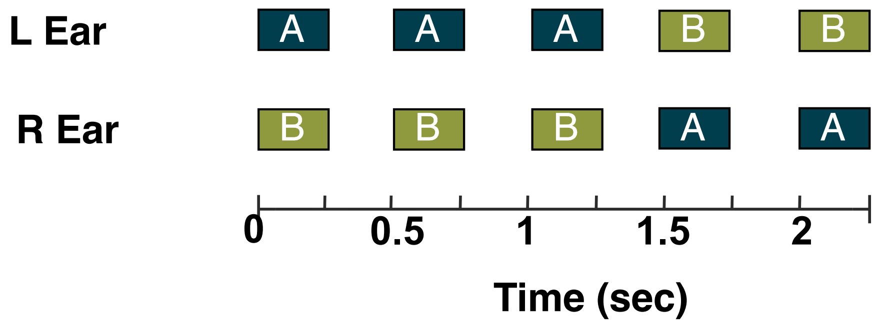


Figure 4.7: *Schematic representation of the stimulus used by Deutsch and Roll (1976) to investigate which low and high tones contributed to the percept of the octave illusion. Blue boxes marked as 'A' indicate low frequency tones (400 Hz) whereas green boxes marked as 'B' indicate high frequency tones (800 Hz). Adapted from Deutsch and Roll, 1976.*

However, their paradigm has several pitfalls including the crucial fact that the stimulus design does not tap into the exact illusion itself. Furthermore, despite having an extremely large sample size, the effect size supporting the Deutsch model (that suggesting that the sequence of pitch in only one of the ears was heard) was extremely small.

The results from the two psychoacoustic paradigms (amplitude modulated tones as well as faded tones) suggest that the concurrent tones may be the ones responsible for the octave illusion percept. These psychophysics results need to be further explored as well as confirmed using neural correlates in order to confirm whether this alternative hypothesis holds true. Chapter 5 builds on this hypothesis by using EEG to confirm whether the concurrent tones are indeed the ones contributing to the percept. Chapter 5 also uses an objective psychophysics paradigm using a deviant detection task to ensure that the listeners are perceiving the illusion in a particular configuration based on the previously established priming tone paradigm (used in Chapter 3).

Lastly, if one considers that the coherent tones are the ones contributing to the pitch percept of the illusion, it makes this stimulus interesting from the point of view of the interaction between concurrent and sequential sound segregation.

Chapter 5

Summary

This chapter describes an EEG study used to further understand the mechanisms of the octave illusion percept. The stimulus used to elicit the octave illusion has low and high frequency tones in both ears; however, the illusory percept arising from it consists of hearing low frequency tones at half the rate of the physical stimulus in one ear and high frequency tones at half the rate of the physical stimulus in the opposite ear (see Figure 3.1 – Chapter 3). In brief, one of the high and low frequency tones each gets suppressed during the percept of the illusion. According to the Deutsch model for the octave illusion (Deutsch, 1981), all the tones presented in one of the ears are enhanced and are heard as the percept of the illusion while all the tones presented in the other ear are suppressed (see Figure 4.1 – Chapter 4). Chapter 4 described two psychophysics paradigms that were aimed at investigating which high and low tones were enhanced and which high and low tones were suppressed. The results from Chapter 4 suggested that contrary to the explanation suggested by Deutsch (1981), the high and low tones heard were the concurrent tone pairs in the stimulus. The current experiment builds on the results of the psychophysics results described in Chapter 4 using EEG to investigate whether the neural activity associated with the octave illusion agrees with

the psychophysical findings, which suggest that the concurrent tone pairs across ears are responsible for the percept. To answer this question of which low and high tones contribute to the illusory percept, the tone streams were differentially tagged using frequency modulation rates that can be seen in the evoked response. The results of the current study support the findings of Chapter 4 and suggest that indeed, the illusion seems to be generated by the concurrent tone pairs.

5.1 Introduction

Chapter 4 used psychophysical measures to support the hypothesis that the components of the concurrent tone pairs are responsible for the illusory percept of the octave illusion. This is an alternative view on the mechanism of the octave illusion, which was most commonly explained by the dual mechanism model suggested by Deutsch (explained in detail in Chapter 4). According to the Deutsch model for the octave illusion (Deutsch, 1981), the tonal percept of the illusion arises from the low and high tones presented in any one of the two ears while the tones of the opposite ear are suppressed (see Figure 4.2 – Chapter 4). This chapter focuses on an EEG study carried out to address the question raised in Chapter 4 regarding the mechanism of the octave illusion. In particular, the aim of this study was to investigate whether the neural activity elicited by the octave illusion supported the hypothesis determined from the psychophysics results of Chapter 4 relating to the synchronous tone pairs contributing to the percept of the octave illusion. A few human neuroimaging studies have been carried out to study various aspects of the octave illusion (briefly mentioned in Chapter 1 section 1.5.3). The studies are described in further detail in this introduction.

Lamminmäki and Hari (2000) carried out one of the first studies to understand the neural basis of the octave illusion using MEG. This study aimed at finding the neurophysiological basis of the 'where' mechanism of Deutsch's dual-mechanism model. The stimuli were 400 and 800 Hz pure tones presented to the left (L) or right (R) ears as follows: L400/R400, L400/R800, L800/R400 and L800/R800. The presentation of these four different stimuli sets was randomized throughout the experiment. The aim of this study was to find out whether the lateralisation of the auditory evoked potentials, in particular the N100m peak, co-varied with the sound localisation percept. The measure used to study the octave illusion was the relative lateralisation of the N100m peak that was assumed to follow the pattern of localisation predicted by Deutsch's model. The original explanation for the octave illusion was that the dominant ear exercises a steady suppression on the other, so that only the frequencies arriving at the dominant ear were heard when the tones were an octave apart from each other (Deutsch and Roll, 1976). In the study by Lamminmäki and Hari (2000), lateralisation of N100m seemed to accompany the perceived sound locations in the octave illusion. The reasoning of the authors was that it was well established that monaural sounds evoke stronger N100m responses in the hemisphere contralateral than that ipsilateral to the sound (Hari, 1990) and participants localised single sound sources to the ear contralateral to the more strongly responding hemisphere. Consequently, they found that the N100m was relatively stronger in the hemisphere contralateral to the high-pitch sound. The authors (Lamminmäki and Hari, 2000) also briefly mention the possibility that stream segregation may be responsible for the lateralisation of the tones, however they do not follow that line of explanation in the study. The drawbacks of this study, in brief, mainly pertain to the fact that this was a passive study where there was no objective behavioural task or even a subjective measure obtained at a trial-by-trial basis to even have a rough measure of what the participants perceived while doing the task. A larger problem with the study is that the EEG was not carried out on the stimulus eliciting the octave illusion itself. The stimuli used were designed to mimic the illusory percept rather than the stimulus that elicited the percept. Using EEG responses evoked by a stimulus that is supposed to mimic the percept rather than the actual illusion-inducing stimuli have the large drawback of presuming what the final percept is supposed to be. Also, a relative lateralisation of averaged N100m peak amplitudes is not a very meaningful way of depicting how the octave illusion may be perceived considering that it is known that the octave illusion is not extremely stable over time and that the percept of the illusion may change in between the trial based on our previous findings

Another method used to study the evoked response in EEG and MEG to the octave illusion is frequency tagging. Frequency tagging refers to a method in which the stimulus (either tones or speech) used for the EEG or MEG experiments are modulated by a specific modulation frequency, which acts like a marker in the overall evoked response. In the case of any stimuli where several components are presented in either an overlapping or synchronous manner, it is difficult to parse apart the evoked responses for every individual component. In such scenarios, temporally overlapping stimuli can be modulated by different frequencies, which can then be used as distinctive markers for individual stimulus components. Several studies have used the method of tagging by different frequency-modulation rates (Bharadwaj et al., 2014; Kaneko et al., 2003; Lamminmäki et al., 2012) to elucidate various overlapping auditory phenomena. The modulation frequencies used in most of the experiments that use frequency tagging (as a method of separating out evoked responses) are preferably frequencies close to 40 Hz. This is because it has been established that the cortical evoked response to stimuli in humans shows maximum amplitude at 40 Hz (Galambos et al., 1981; Hari et al., 1989).

More recently, Bharadwaj et al. (2014) have also used the technique of frequency tagging in an EEG study to investigate selective attention to speech streams comprising of different vowel sets that were presented at the same time but separated in location by differentially tagging the competing streams. The study indicated that frequency tagging is an effective method of studying attention effects on concurrently presented auditory stimuli. They showed this by showing an increase in amplitude of the tagged frequency for the attended stream of vowels. The study also used tagging frequencies equidistant from 40 Hz (approximately 35 and 45 Hz), which are the frequencies used in the current study as well.

These studies illustrate that frequency tagging can be a useful tool to tease apart parts of an overlapping stimuli. Additionally, an increase in amplitude for a single tagged frequency can be seen in the salient stimuli (Bharadwaj et al., 2014), even within concurrent sound sequences.

Lamminmäki et al. (2012) used the method of frequency tagging in another MEG study that aimed to explore the link between the perceived pitches and interactions of dichotic tones in the octave illusion. Their stimuli were eight combinations of 400-Hz and 800-Hz pure tones. There were four monaural stimuli, L400, R400, L800, and R800, (where L refers to left-ear input, R to right-ear input, and the number to the tone frequency - either 400 or 800 Hz), two binaural same-pitch stimuli (diotic), L400R400 and L800R800, where either the 400-Hz or the 800-Hz tone was presented to both ears, and two dichotic stimuli, L400R800 and L800R400, which they suggested were the “octave-illusion stimuli”. Furthermore, their stimulus was tagged with differential modulation frequencies; the LE inputs were modulated at 41.1Hz and the RE inputs at 39.1 Hz. This study aimed at finding the neurophysiological basis of the ‘what’ mechanism of Deutsch’s dual-mechanism model (the ‘what’ mechanism mainly being the pitch of the tones perceived in the octave illusion). The results were based on the amplitude of the ipsilateral and contralateral frequency-tagged MEG responses. They found that the strength of the MEG activity to the tagged frequencies of the tones varied depending on whether a tone of the same frequency or a tone separated by an octave was presented simultaneously to the other ear. They attributed this difference to modified binaural interaction. They further found that there was an increased right ear dominance for all their participants (all participants used were right handed) which they suggested was in line with Deutsch’s model that suggests that the pitch percept comes from the tones in the dominant ear. The authors suggest that this modified binaural interaction where contralateral non-dominant hemispheric signals were suppressed could be deemed as the neural correlates to the ‘what’ pathway suggested by Deutsch (1975). This study suffers from the same drawbacks as the previous study by Lamminmäki and Hari (2000) where all the findings are based on stimuli that reflect the reported percept from the illusion rather than the actual octave illusion stimulus. It must be noted that this study was also carried out in a passive listening condition, where the listeners' percept was not controlled.

The current study uses the stimulus that is similar to the dichotic stimulus used to elicit the octave illusion (Deutsch, 1974) as shown in Figure 5.1 in contrast to the studies described above that tend to use a stimulus that replicates the illusory percept. In this study, the tones of the stimulus were differentially tagged to obtain a direct measure of which tones were perceived during the illusion.

The aim of the current study was to build on the results of Chapter 4 that indicated that the concurrent tones were the tones that were responsible for the pitch percept of the octave illusion. The hypothesis of this study was that the modulation frequency corresponding to the contralateral high tones concurrent with the attended low tones would show an increase in amplitude. For example, if the listener has a percept of Right ear – Low tones and Left Ear – High tones, the amplitude of the modulation frequency of the concurrent high tones in the Left ear would be larger than the amplitude of the modulation frequency of the high tones in the right ear.

5.2 Methods

5.2.1 Participants

Thirteen adults (21–30 years of age, 6 male) with normal hearing (audiometric thresholds under 20 dBSPL for octave frequencies between 250 Hz and 8000 Hz) and reportedly no history of neurological disorders participated in the experiment. Informed consent was obtained after the procedures of the experiment were explained to them. They received payment for their participation. The University College London Ethics Committee approved the experiment. All participants were naïve to the stimulus and were chosen on the basis of their ability to perceive the octave illusion without any precursors. A short behavioural block of listening to the unbiased octave illusion test sequence and noting the subjective responses assessed this. All participants could perceive the octave illusion (i.e. low tones in one ear and high tones in the opposite ear).

5.2.2 Stimulus and Paradigm

The stimulus paradigm is similar to the ones used in Chapters 3 and 4. A schematic of an example trial is shown in Figure 5.1. At the start of each trial, a precursor consisting of three low tones at a frequency of 1000 Hz was either presented in the left or right ear. Each tone was 200 ms long with a silent gap of 100 ms between the three tones. The sequence of 3 tones was followed by a 1000 ms silent gap. Following the 1000 ms silence, the test sequence was played. In the test sequence, each ear was presented with a sequence of alternating high and low tones. The alternating tone sequences in each ear were one tone out of phase (for example, If the left ear sequence started as High-Low-High, the right ear sequence started as Low-High-Low). In Figure 5.1, the low tones (frequency = 1000 Hz) are indicated as the boxes marked ‘A’ and the high tones (frequency = 2998 Hz) are marked ‘B’. The high tones in each ear were differentially modulated by modulation frequencies of either 34.4615 Hz or 44.3077 Hz. Each tone was 203.1 ms long followed by a 50 ms silent gap.

The tone lengths and precise frequencies were adjusted to these values so as to ensure that each high frequency tone had an integer number of complete cycles of both the modulation frequencies as well as ensure that the modulation envelope was at zero at the start and end of the stimulus. This was done in order to minimize splatter during analysis (section 5.3). Each test sequence consisted of 40 tones. The total duration of the test sequence was 9950 ms. The listeners had to carry out a target deviant detection task (similar to the task in Chapter 3). However, the deviants were present only on the low frequency tones. The listeners had to attend to the low frequency tone stream in the ear cued by the precursor sequence and indicate whether an increase in amplitude was present on one of the target stream low tones in the attended stream while ignoring any other distractor deviants. For example, in Figure 5.1, the precursor sequence is made of low frequency tones in the right ear. The corresponding percept of this trial would be low frequency tones in the right ear and high frequency tones in the left ear. The listeners carried out a deviant detection task where the deviant in the right ear would be the target deviant whereas the deviant in the left ear would be the distractor deviant. 50% of the trials had target deviants whereas all the trials had one or more distractor deviants. In each trial containing a target deviant, there was only one target tone whereas the number of distractor tones in the opposite ear varied from trial to trial. The sequences were generated in MATLAB (MathWorks Inc. Natick, MA, USA) and were presented at a sampling rate of 44.1 kHz.

The total EEG stimulus set was counterbalanced for probe type v/s tagging frequency. The four stimulus conditions were:

1.Probe: Left ear.

Right ear tagging frequency: 44.31 Hz

Left ear tagging frequency: 34.47 Hz

2.Probe: Right ear.

Right ear tagging frequency: 44.31 Hz

Left ear tagging frequency: 34.47 Hz

3.Probe: Left ear.

Right ear tagging frequency: 34.47 Hz

Left ear tagging frequency: 44.31 Hz

4.Probe: Right ear.

Right ear tagging frequency: 34.47 Hz

Left ear tagging frequency: 44.31 Hz

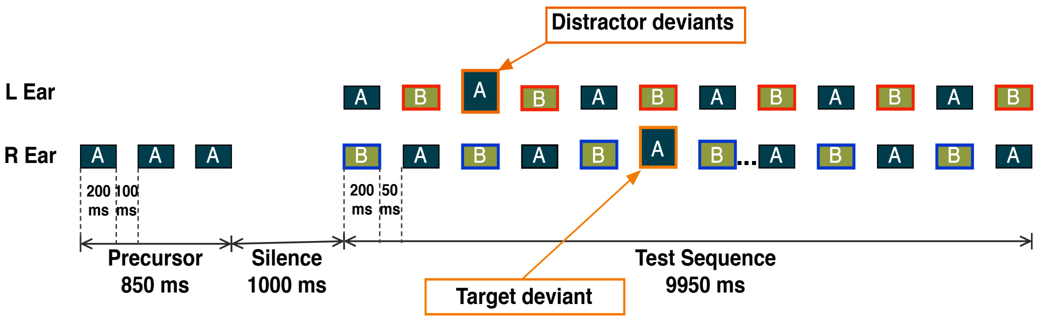


Figure 5.1: *Test stimuli example. Each ear was presented with opposing, alternating frequency sequences of pure tones (A = 1000 Hz with no modulation; B = 2996 Hz tagged with modulation frequencies of 34.4 Hz or 44.3 Hz). Listeners were cued to focus on the low frequency precursor on either side indicated by a priming sequence and were asked to detect target amplitude deviants. The schematic diagram below shows a sample trial where the right ear and left ear high tones are differentially tagged (red and blue outlines) and the low frequency tone cues are in the right ear.*

Two control conditions were also added to the EEG paradigm to establish a baseline for the tagged frequencies. An example of the control condition stimulus is shown in Figure 5.2. As shown, this stimulus is not alternating and hence does not elicit the octave illusion. The stimuli had all low tones in one ear and all high tones presented concurrently in the opposite ear. The high tones were tagged with either of the two modulation frequencies.

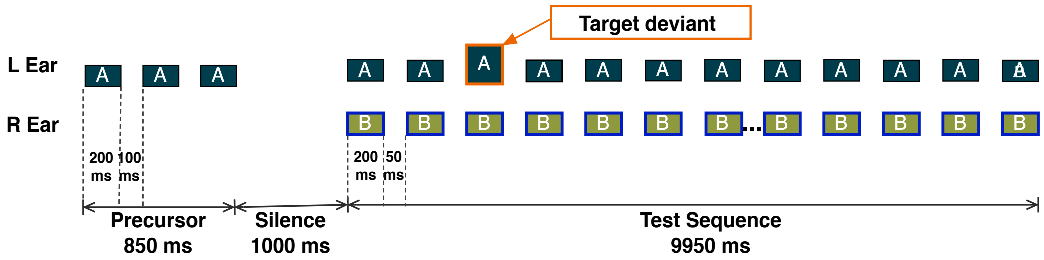


Figure 5.2: *Control stimuli example. Each ear was presented with single frequency sequences of pure tones (A = 1000 Hz with no modulation; B = 2996 Hz tagged with modulation frequencies of 34.4 Hz or 44.3 Hz). Listeners were cued to focus on the low frequency precursor on either side indicated by a priming sequence and were asked to detect target amplitude deviants. The schematic diagram below shows a sample trial where the right ear high tones are tagged (blue outlines) and the low frequency tone cues are in the left ear. This stimulus paradigm does not elicit the illusory percept.*

5.2.3 Procedure

The first part of the experiment was a behavioural set where the listeners first heard the unbiased illusory sequence with only pure tones (i.e. no precursor tones and no modulation). Their unbiased percepts were recorded for ten repetitions of the stimulus. Next, they carried out another set of ten repetitions where their perceptual responses to the experimental test stimulus without the precursors used in the experiment (low frequency pure tones with high frequency modulated tones – as described in section 5.2.2) were noted. Lastly, they carried out a set of behavioural trials where their subjective responses to the sequences with a biasing precursor sequence were noted. The subjective responses were noted via button press to record the percept of the octave illusion for each precursor sequence (either Right High-Left Low or vice versa). This last set of trials had ten randomized repetitions of each precursor sequence. For all the three behavioural sets described above, the listeners were all naïve to the stimuli and were not told what the expected response for the octave illusion was.

The second part of the experiment involved recording the EEG responses to the stimuli depicted in Figs. 5.1 and 5.2. EEG was acquired continuously using a 64-channel BioSemi active-electrode EEG system (Biosemi Inc., Amsterdam, Netherlands). The signal was digitally sampled at an A/D rate of 2048 Hz (64-bit resolution). Participants were fitted with an electrode cap fitted with 64 silver/silver chloride scalp electrodes. Electrode impedance was monitored and maintained at a minimum (typically below 5 kΩ). The stimulus was presented in blocks that were either ‘test’ blocks or ‘control’ blocks. Within each of the blocks, the trials were randomized for probe type as well as tagging frequency. Each block consisted of 240 trials and each listener carried out 3 test blocks and 2 control blocks. For each trial, the listeners had to focus on the cued stream (as determined by the precursor). At the end of each trial, the listener had to report if a target deviant was heard or not with a button press. The next trial was initiated with a 1 second time delay post response.

5.3 Analysis and Results

5.3.1 EEG pre-processing

EEG pre-processing, epoching and averaging was carried out using the EEGLAB toolbox (Delorme and Makeig, 2004). Data was filtered using a zero-phase-shift bandpass filter from 0.1 Hz to 70 Hz. Baseline was corrected to -500 ms, followed by artefact rejection at +/- 150 microvolts. Independent component analysis (ICA) was used to remove artefacts related to eye movements and blinks. All electrodes were included for the ICA analysis. The data was epoched according to the six conditions described in section 5.2.2 (four test and two control) as each condition had a different EEG trigger. The EEG data further analysed was averaged across a select subset of channels from the left, right and central electrode positions over the temporal and parietal regions (FC3, C1, TP7, P5, P7, P9, FPz, FC4, FC2, FCz, Cz, C2, P6, P8, P10). The data was analysed in terms of relative spectral strength of the tagged frequencies across conditions as well as analysed for differences in the EEG waveform itself (shown in the period histograms in section 5.3.3). The data was averaged over all the 15 electrodes for these analyses.

5.3.2 Frequency analysis

The EEG test stimuli (Figure 5.1) were designed in a way that the two high-frequency tone streams were tagged with different modulation frequencies. This was to tease apart whether the coherent tones or the sequential tones were the tones that contributed to the percept of the octave illusion. The part of the EEG signal that was analysed was only the EEG response corresponding to the test sequence. Hence, the first step was to epoch the EEG signal from the onset of the test sequence till the end of the test sequence thereby cutting out the EEG related to the precursor, the silent period in between as well as the motor response at the end of the trial (9950 ms in length (see Figure 5.1). Furthermore, the response of the first and last tone pairs was also excluded in order to exclude the onset and offset responses. Next, the EEG signal only to the exact length of each of the tones (corresponding to the 203.1 ms long tones) was extracted and concatenated. The time points for the start and end of each tone were marked and extracted and the EEG for each of these 40 tones was then concatenated. This was essential to get minimal splatter during the Fourier transform. Since the stimulus was designed so that the start and end phase of the modulation envelope of each tone was zero, on concatenation, no phase differences were present. A Fast Fourier Transform was then carried out on this concatenated sequence of the EEG only related to the tones. Figure 5.3 indicates the peaks at the tagged frequencies across all test conditions.

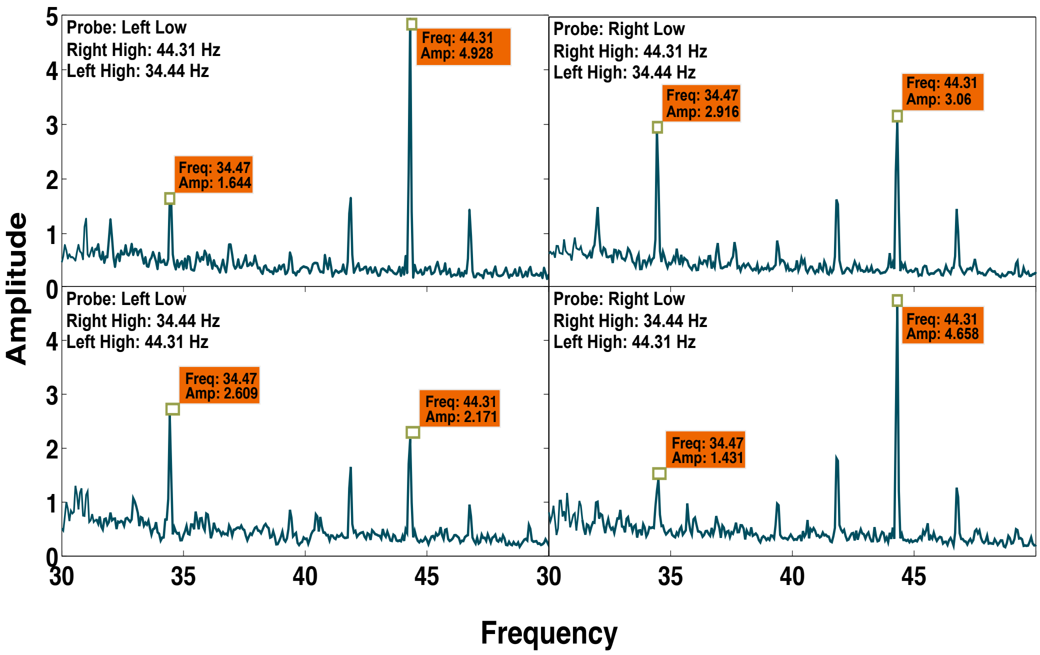


Figure 5.3: *Spectrum of tagged frequencies for the test sequences. The figure shows the raw spectrums of the test signals for all test sequences. In all test sequences, the tagged frequency corresponding to the concurrent tone is amplified.*

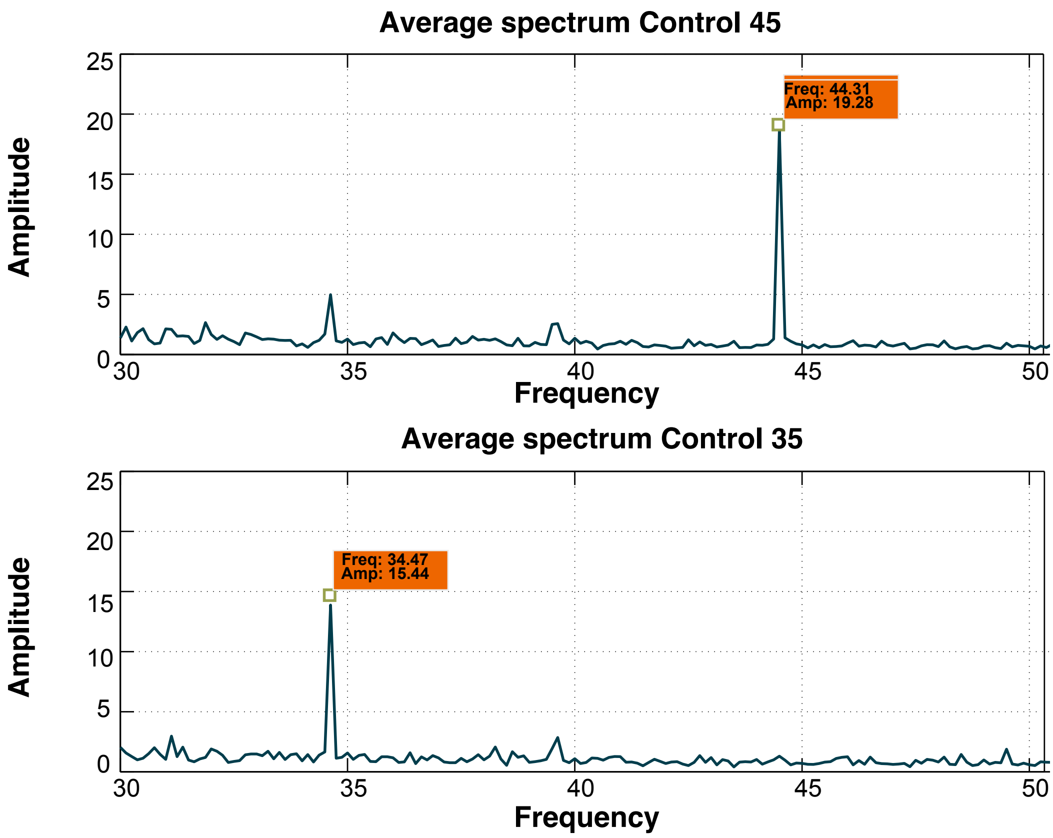


Figure 5.4: *Spectrum of tagged frequencies for the control sequences. The figure shows the raw spectrums of the test signals both control sequences as a baseline measure. The figures clearly indicate that the tone at 44.31 Hz evokes a larger EEG signal than the tone at 34.47 Hz.*

As seen in Figure 5.4, the amplitudes of the tagged frequencies in the control condition are different (higher for the tagged frequency of 44.31 Hz). In order to get a better idea of the changes in tagged frequency amplitudes in the test conditions, the ratio of the peak amplitudes at the tagging frequencies was measured for each subject. The ratio shown in Figure 5.5 is the ratio of the amplitude of the 44.31 Hz peak to the amplitude of the 34.47 Hz peak. There was a statistically significant effect of frequency condition, $F(3, 48) = 13.84$, $p<0.005$. Post-hoc comparisons indicate a significant difference between conditions ($p<0.005$) as shown in Figure 5.5. The condition with the precursor sequence in the left ear where the concurrent high tone in the Right ear was tagged at 44.31 Hz differed significantly (p<0.005) from Conditions 2 and 3 where the concurrent high tone was tagged with 34.47 Hz. Condition 4 which also had the concurrent tone tagged at 44.31 Hz differed significantly from Condition 3 (p<0.01). However, the difference between the two conditions with the precursor as Right-Low trended towards significance (see Figure 5.5) but was not statistically significant. In all four conditions, the tone stream that was coherent to the probe tone showed higher amplitude of the tagged frequency.

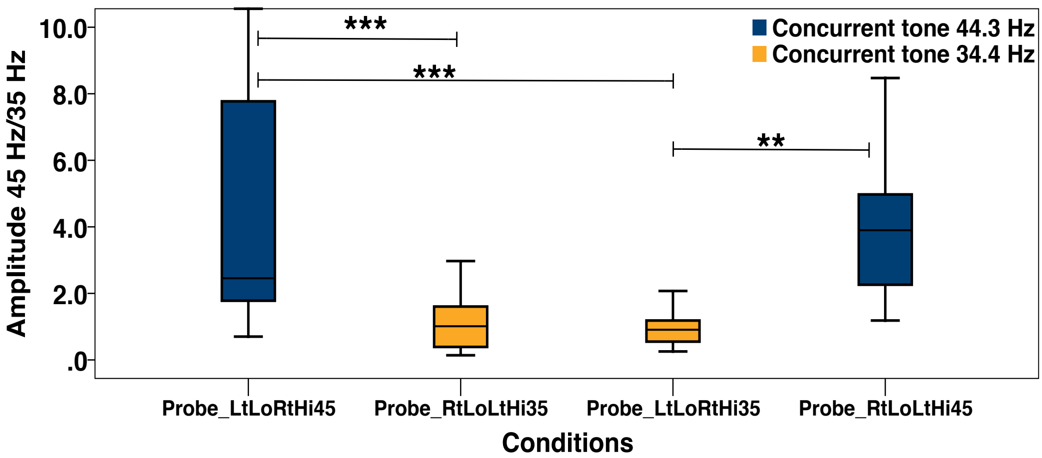
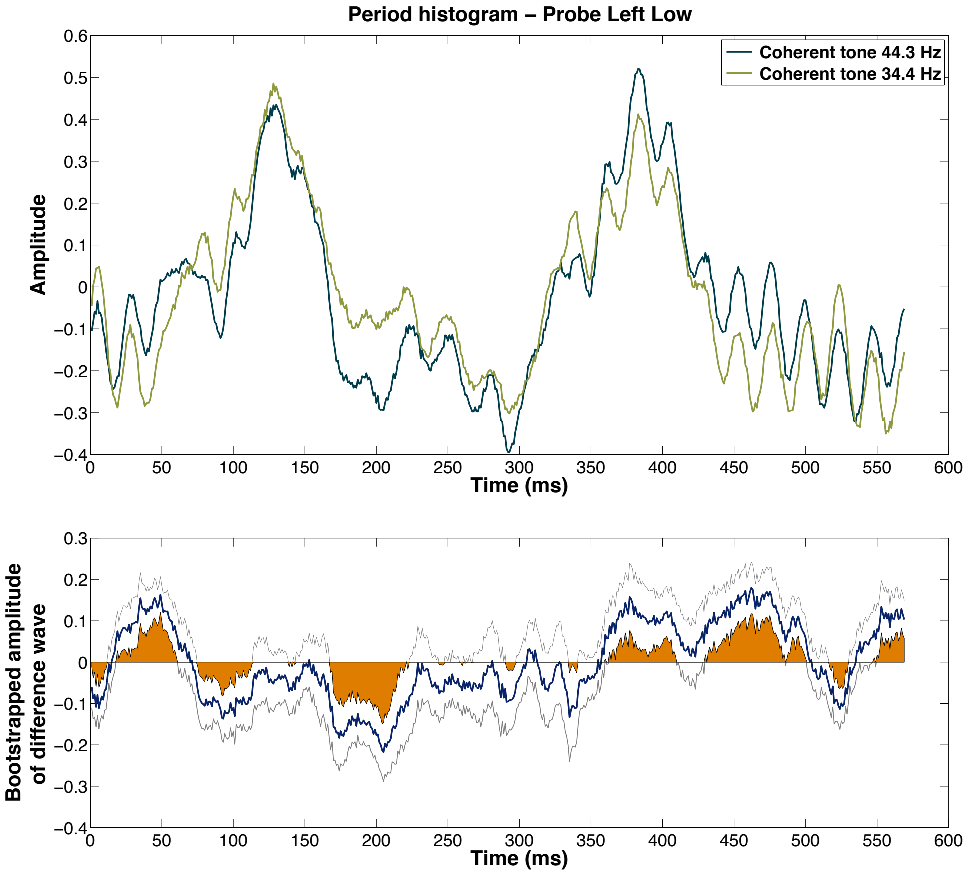


Figure 5.5: *The amplitudes of the tagged frequencies for each test condition were measured as a ratio of the amplitudes of response components at the two rates of 44.31 and 34.47 Hz. In conditions where the concurrent tone was tagged with 44.31 Hz, the ratio is found to be significantly higher than the conditions where the concurrent tone was tagged with 34.47 Hz.*

5.3.3 Period histogram analysis

Apart from the frequency analysis of the data, the averaged EEG waveforms from the subset of electrodes (see section 5.3.1) for each participant were also compared to help visualize the difference between test conditions better. The data corresponding to the test sequence was segmented into portions consisting of the EEG response to two concurrent tone pairs (i.e. segments of two tones with two silent periods). These segments were then averaged for each condition to form a period histogram for each of the four conditions. As seen from Figure 5.6, the modulation of the tones can be clearly visualized as the faster modulations in the EEG waveform. In both panels, the first concurrent tone pair has a high frequency tone modulated at 34.47 Hz and the second concurrent tone pair has a high frequency tone modulated at 44.31 Hz (seen as a faster modulation). The level of significant difference between the conditions across the time period of the period histogram was evaluated using the bootstrap method (Efron and Tibshirani, 1993), using 1000 bootstrap iterations, based on the individual difference waves. For illustration purposes, the group grand average is plotted in Figure 5.6. For illustration purposes, group averages are plotted, but statistical analysis was always performed across participants.

To compare the EEG activity between conditions for the same probe (for e.g. Panel 1 in Figure 5.6 shows the difference between the two different tagged conditions where the probe was Left Low), a repeated-measures analysis was used in which, for each participant, the average time series for one condition was subtracted from the average time series for the other condition, and individual time-series differences were subjected to bootstrap analysis with 1000 iterations (Efron & Tibshirani, 1993). At each time point, the proportion of iterations below the zero line was counted (see bottom of panels in Figure 5.6). If the proportion was less than 0.01% or more than 99.99% for 10 adjacent samples, the difference was judged to be significant. The shaded area (in orange) indicates the temporal interval over which a repeated-measures bootstrap procedure revealed a significant difference between responses to the two stimuli. A significant difference was found for the period histograms for the conditions with the low frequency precursor sequence in the left ear. The difference was not as significant for the other two conditions where the precursor was in the right ear.


Period histogram – Probe Left Low
Coherent tone 44.3 Hz
Coherent tone 34.4 Hz
Amplitude
Time (ms)
Bootstrapped amplitude of difference wave
Time (ms)

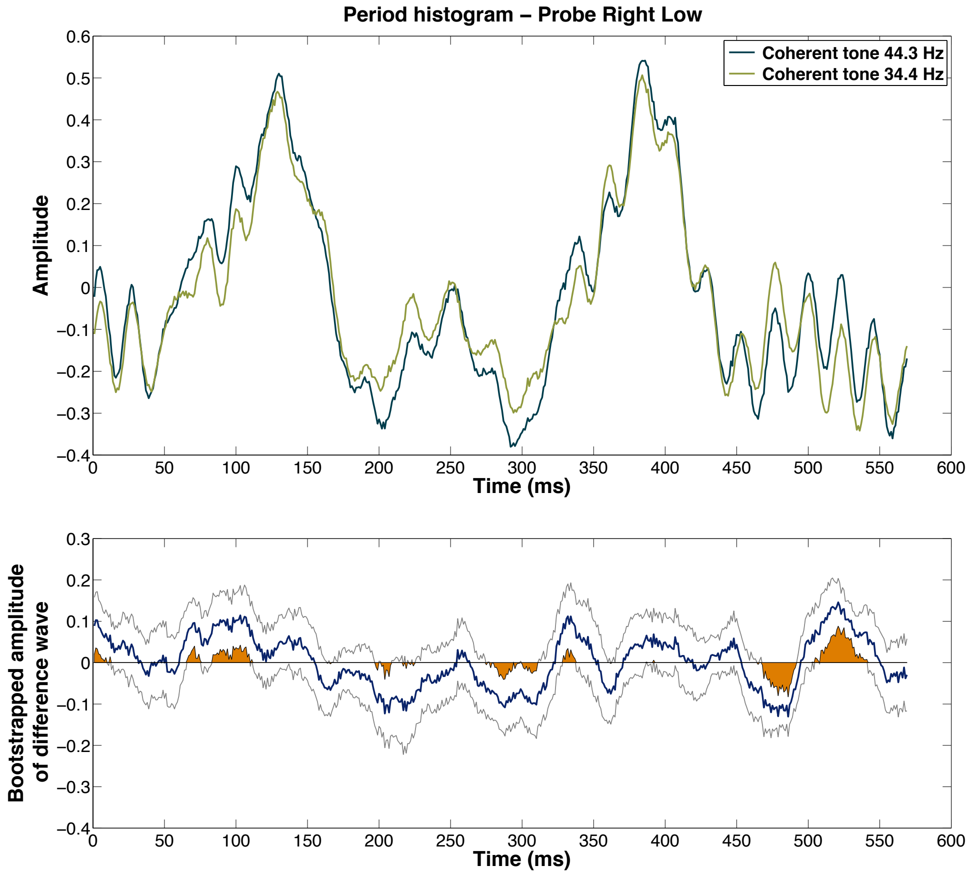


Figure 5.6: *Period histograms for each probe condition. Upper two panels show the raw waveform period histograms and bootstrap results of the difference waves for the probe in the left ear (Left Low) and the lower panels show the raw waveform period histograms and bootstrap results of the difference waves for the probe in the right ear (Right Low). At each time point, the difference is deemed to be significant if the proportion of iterations above/below the zero line is less than 0.005 for 5 adjacent samples. The blue lines in each period histogram denotes the EEG activity for conditions where the coherent tone was 44.3 Hz whereas the green lines in each period histogram denotes the EEG activity for conditions where the coherent tone was 34.4 Hz. In the bootstrap plots, the blue line indicates the difference wave and the orange shaded areas indicate the significance areas.*

5.4 Discussion

The aim of this experiment was to confirm the findings of Chapter 4 using EEG. The results of chapter 4 indicated that the octave illusion of alternating tones seemed to arise from concurrent, rather than sequential tones, suggesting that the octave illusion arose from a misattribution of time across perceptual streams, rather than a misattribution of location within a stream. The results from the current EEG study also indicate that indeed the concurrent tones are perceived as most salient and are amplified. Previous EEG studies related to investigating the octave illusion have found results that partially support the Deutsch dual-mechanism model (Lamminmäki and Hari, 2000; 2012). However, all of these studies did not objectively control for the attention of the listeners. The percept elicited by the octave illusion is variable and can be manipulated by directed attention as shown in the results of Chapter 3. By using the method of a precursor sequence combined with an objective deviant-detection task (used in both Chapter 3 as well as in the current experiment), there is a higher degree of control over understanding what tones are salient to the listener. The stimulus related to the octave illusion generates a salient percept of either Right – Low Frequency alternating with Left – High Frequency or vice versa. In all subjective reports with and without directed attention, all listeners always perceived two salient tones. Hence, it was hypothesised that on every trial, two tones were predominantly salient: the low frequency tone corresponding to the precursor cue and a high frequency tone heard asynchronously to the low frequency tone in the opposite ear.

The EEG results in the study described in this chapter indicated that for every precursor condition, the power of the tagged frequency, i.e. the amplitude of the evoked response pertaining to the particular frequency tag, corresponding to the tone presented concurrently with the target stream was higher. For example, in a trial where the precursor tone of low frequency tones was in the left ear leading to an illusory percept of Left-Low: Right-High tones, the tagged frequencies corresponding to the Right ear high tones was greater. Both tagged frequencies were in every test condition as both frequencies were physically present in the stimulus. However, across probe conditions as well as tagging frequencies, all results indicated that the tone coherent with the cued tone was amplified.

The interesting aspect of this result was that the coherent tones sound asynchronous in this particular stimulus pattern. It has been known that two tones of different frequencies are difficult to parse apart due to the strong binding cues of temporal coherence (Elhilali et al., 2009; Micheyl et al., 2010; Micheyl et al., 2013). However, in this particular illusory setup, each coherent tone pair alternates with its exact inverse coherent pair. In such a scenario, the coherent tone pairs seem to separate into two streams of their individual frequencies. This hypothesis is theoretically explored in the next chapter where the computational neural model of the octave illusion has been described which shows the two coherent tone streams segregate into different channels. Thus, the stimulus in each ear has one salient frequency that groups the similar frequency components in that particular ear as a sequential stream. For example, if the low frequency cue were in the Right ear, the resulting percept would be Right Low-Left High. Furthermore, the low frequencies in the Right ear and the high frequencies in the left ear would segregate from the other frequencies in each ear and the resultant would be two sequential streams of low and high frequency tones in the right and left ears respectively.

It is known that listeners are poor at judging temporal synchrony or asynchrony between two parallel sound streams and that stream segregation has a detrimental effect on the perception of temporal relationships between segregated sounds (Micheyl et al., 2010; Roberts et al., 2002; Vliegen et al., 1999). This would then explain the stream sounding asynchronous between the two ears. A more thorough explanation of this will be provided in the next chapter that explains the model related to the octave illusion.

Chapter 6

Summary

This chapter outlines a new model (developed in collaboration with Shihab Shamma, Nori Jacoby and Andrew Oxenham) developed to explain the octave illusion in light of the findings of the studies described in Chapters 3, 4 and 5. In brief, the results of these studies indicated the following properties of the octave illusion: 1) the octave illusion can be manipulated by selective attention (Chapter 3), 2) stream segregation plays a role in the mechanism of the illusion (Chapter 3) and 3) attended tone pairs concurrent across ears contribute to the tones of the illusory percept that are predominantly perceived (Chapters 4 and 5). The dual-mechanism model currently in place to explain the octave illusion (Deutsch, 1981) does not account for selective attention nor does it support the findings that the concurrent tones cause the pitch percept of the octave illusion. The new model described in this chapter is based on the three main findings of Chapters 3, 4 and 5 (listed above).

6.1 Introduction

The octave illusion, first reported by Deutsch (1974) is a fascinating auditory illusion that provides a unique insight into stream segregation using a relatively simple, two-tone stimulus. The stimulus used to elicit this illusion consists of opposing sequences of two alternating pure tones delivered to each ear of the listener (see Figure 4.1 – Chapter 4). The resulting percept is that of the high frequency and low frequency tones being perceived at half the rate of the stimulus with all the high tones lateralising to one ear and the low tones to the opposite ear (see Figure 4.1 – Chapter 4). Even though at any given point, all tone pairs of low and high tones are temporally coherent, the resulting percept is not one of a fused complex tone (which would typically be the case for temporally coherent tone pairs) (Elhilali et al., 2009; Micheyl and Oxenham, 2010; Micheyl et al., 2013). From a stream segregation point of view, this provides a unique setting of tones where tones that are temporally coherent seem to separate out easily. The dual mechanism model (Deutsch, 1981) used to describe the mechanisms of this illusion has been described in detail in Chapter 4 (section 4.2). However, the data related to the octave illusion from Chapters 3, 4 and 5 cannot be fully explained by Deutsch's dual mechanism model. This chapter highlights how the dual mechanism model fails to explain the findings of the studies described in Chapters 3, 4 and 5 and proposes a simple alternative model (not based on the dual mechanism model suggested by Deutsch, 1981) that helps explain the octave illusion in light of our findings.

6.2 Arguments against the Deutsch dual-mechanism model

The dual-mechanism model (Deutsch, 1983) has been studied and supported over the past thirty years and Deutsch and colleagues have extensively run subjective psychoacoustic tests on the octave illusion (Deutsch, 1978; 1980; 1981; 1987; 1988; Deutsch & Roll, 1976). Further studies including human EEG and MEG studies related to the octave illusion have been carried out and to some extent, these support the Deutsch dual mechanism model (Lamminmäki and Hari, 2000; Lamminmäki et al., 2012; Brancucci et al., 2009; 2011). However, the shortcomings of these studies have been discussed in the previous chapters. In brief, the criticism against the electrophysiological studies using EEG and MEG to study the octave illusion are 1) they often use stimuli that mimics the expected illusory percept and 2) all of these studies are carried out in a passive listening state which has an inherent confound of having no control over the perceptual state of the listener.

Furthermore, the characteristics of the data obtained in the studies described in this thesis also are not in agreement with the model. Chapter 3 indicates that the percept can be manipulated using attention. This indicates that the percept can be heard both as the low tones in the left ear and high tones in the right or vice versa, irrespective of the handedness of the candidate. This firstly does not agree with the model's pitch percept explanation suggesting that the pitch of the tones only relating to the dominant ear are heard. Further, Chapter 4 and 5 indicate that in fact, the tones heard as alternating or asynchronous in time are the concurrent tone pairs. The current model cannot explain how the concurrent tones are the ones contributing to the tonal percept of the illusion. According to the current model, the pitch of the tones arise from either the left or right ear tone sequences. Lastly, the model does not adequately explain how the illusory percept would break down if the tones differed in frequencies (for example, if Right Low and Left Low were different frequency tones).

In order to explain our experimental findings, a preliminary model has been proposed in the next section.

6.3 New proposed model

In order to understand this model, we start by considering two types of frequency-specific neural units, monaural and binaural. In the case of the octave illusion stimulus (as shown in Figure 6.1), four monaural channels (Right ear Low frequency (RLo), Right ear High frequency (RHi), Left ear Low frequency (LLo) and Left ear High frequency (LHi)) and two frequency-specific binaural channels (Binaural High and Binaural Low) can occur. The octave illusion stimulus can also be thought of as two sets of alternating, concurrent tone pairs: X and Y (X = RLo and LHi and Y = RHi and LLo). The tones within X and Y are temporally coherent and if either X or Y were heard in isolation, they would elicit a fused percept of a complex tone.

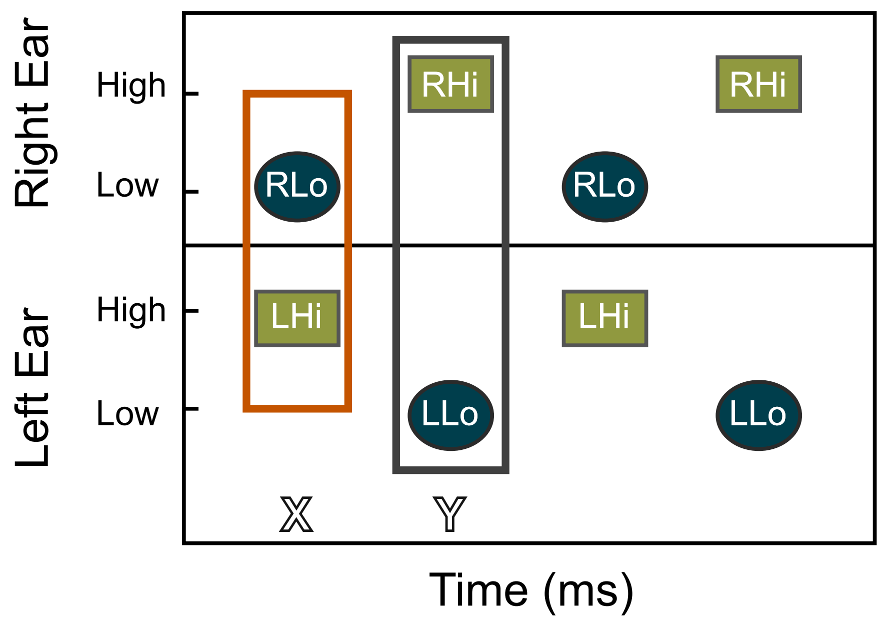


Figure 6.1: *Schematic representation of a snapshot of the stimulus used to elicit the octave illusion. RLo and LLo pure tones have the same low frequency (indicated in dark blue) and LHi and RHi pure tones have the same high frequency (indicated in green). The tones across ears are also temporally coherent. Hence the stimulus can be described in terms of two alternating pairs of coherent tones (indicated by X and Y).*

Given this framework, the model's working can be described in three broad steps or phases as shown in Figure 6.2.

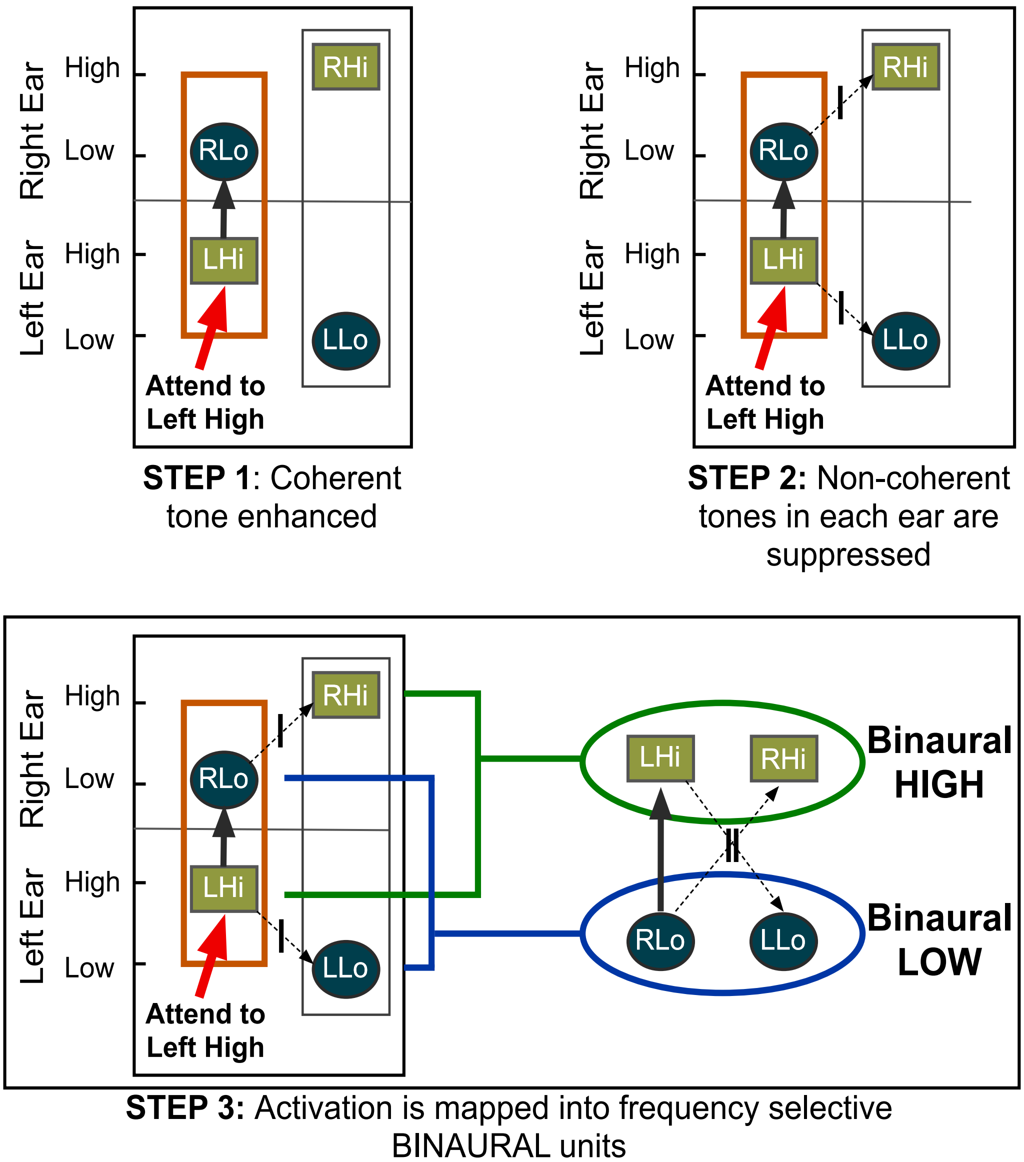


Figure 6.2: *Schematic representation of the proposed model to explain the octave illusion. See section 6.3 for a detailed description.*

<u>Step 1:</u> Consider a scenario where a listener is cued to attend to the Left ear High tone (LHi). In this case, the tone coherent with the attended tone

(in this case, Right ear Low tone (RLo)) is enhanced by virtue of being temporally coherent with the attended tone. As a result of this, X (consisting of concurrent tones LHi and RLo) is enhanced relative to Y (consisting of concurrent tones LLo and RHi) (Figure 6.2 - Step 1).

Step 2: In each ear, a low and high tone is present. Due to attention, one tone in each ear was enhanced (in the current example, RLo in the right ear and LHi in the left ear). In the case of the octave illusion, each ear can also be seen as a monaural, two-tone sequential streaming paradigm. So in order to stream themselves apart, the enhanced tones suppress the competing monaural tones. In the current example, the RLo suppresses the RHi in the right ear and the LHi suppresses the LLo in the left ear (Figure 6.2 - Step 2).

Step 3: Now, consider the frequency specific binaural units. These are activated at all times by the low tones and high tones *across* ears. The RLo and LLo activate the Binaural Low units and the RHi and the LHi activate the Binaural High units. However, as indicated in Figure 6.2, the within-ear competition/suppression between the tones (LHi vs. LLo and RLo vs. RHi) now translates into across-unit competition between the binaural units (Binaural High and Binaural Low). As a result of this across-unit suppression/competition, the binaural units are no longer correlated and can hence be split into separate channels of different frequency. This in turn results in the coherent X tones being grouped into separate, competing binaural units (Figure 6.2 - Step 3). Hence, although the tones heard most saliently will be the tones of the complex, X (RLo and LHi), they no longer group together as a coherent whole. Instead they will be perceived as two independent tone streams in each ear that are no longer perceived as a coherent single stream. It is known that it is difficult to judge the temporal properties of tones that fall into different streams (Micheyl et al., 2010). Hence, the two streams of low and high tones (in the case of this example, RLo and LHi) are heard temporally asynchronous.

The model described above would only work for stimuli, which have a configuration similar to that of the octave illusion stimulus. This would mean that X and Y would need to consist of the same two frequencies to enable the low tones and high tones to be grouped into the same binaural units. Due to this, the current model also explains why the illusion breaks down if the tones in Y differ in frequency from the tones in X. This can be seen in Figure 6.3 where all four tones have different frequencies and hence belong to different binaural units. In the case of Figure 6.3, when the listener attends to LHi (Left ear Hi tones), the model depicts how the four tones would fall into separate binaural cells. Notice that there is no competition between the binaural units containing the attended coherent tones, hence one would not expect them to sound segregated and instead they sound as a coherent tone unit.

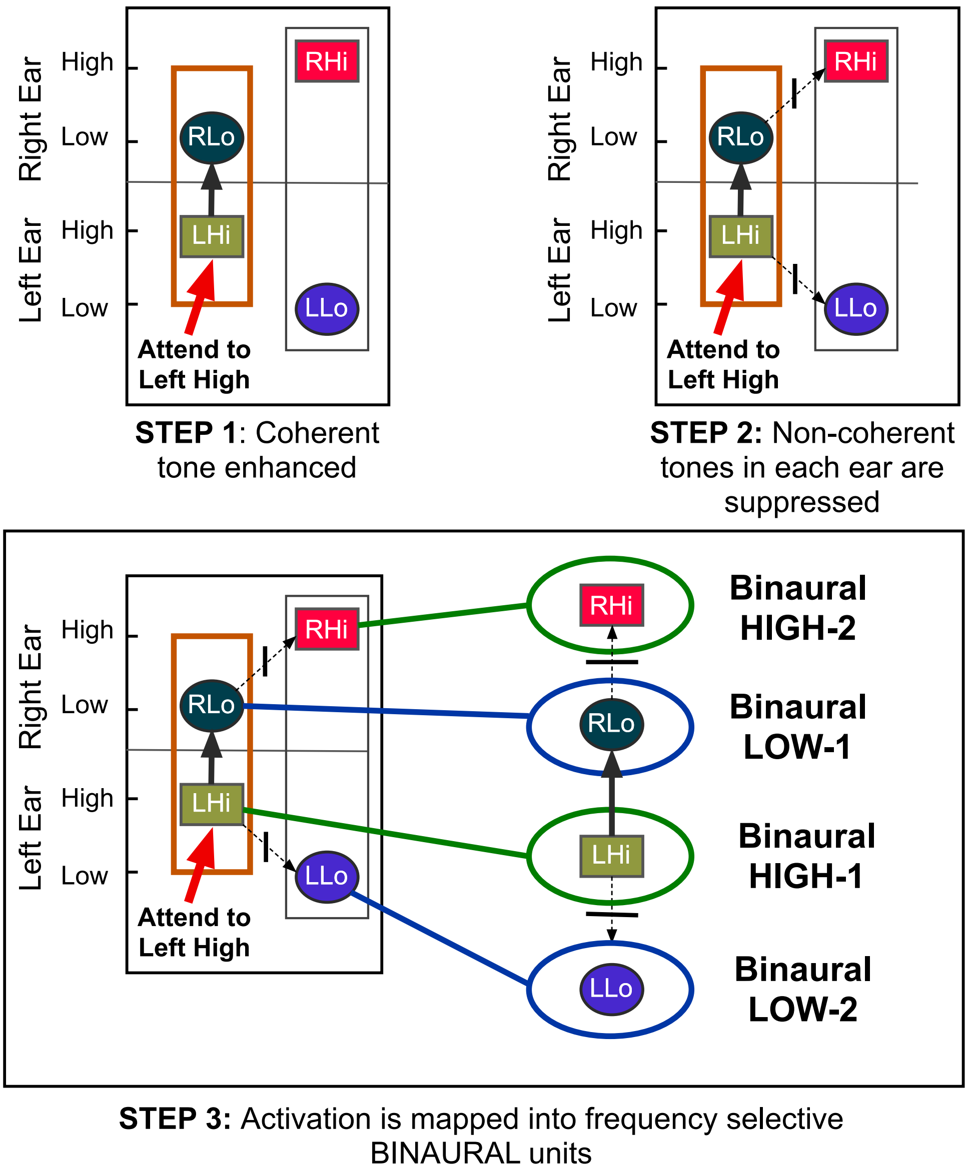


Figure 6.3: *Schematic diagram showing how the illusion would not occur if the frequencies of the four component tones were different. See section 6.3 for details.*

Lastly, the new model also explains how manipulating attention can change the percept. The percept can be varied from RLo-LHi to RHi-LLo by simply attending to any of the different component tones in either ear. This model explains the results of Chapters 3, 4 and 5, which suggest that the illusion can be perceived in any fashion with a suitable priming sequence and that the illusion arises from the synchronously presented tones in the stimulus.

6.4 Discussion

The new model described here is well supported by the current views in the field of stream segregation. Temporal coherence is known to be an important binding feature for dichotic stimuli and in the octave illusion, a specific feature of coherence has been highlighted. The proposed model suggests the increase in saliency of the coherent tone pairs that are attended to by listeners, which again is a well-established phenomenon (Fritz et al., 2007). Furthermore, within ear stream segregation of tones with different frequencies where the responses to one of the tones is amplified is also a fairly established finding (Gutschalk et al., 2005; Micheyl et al., 2005). Lastly, the presence of binaural neurons that are tuned to specific frequencies has been studied in several mammals (Middlebrooks et al., 1980; Joris & Yin, 2007; Razak, 2011). The current model proposes a sequential analysis carried out during the perception of these illusory stimuli and suggests that the competition between the binaural cells leads to the breakdown of the coherence between the two attended tones (as explained in section 6.3). The sequential aspect would also explain the time taken for the illusory percept to build up. This build-up has been reported by Deutsch & Gregory (1978) as well as reflected in the data in chapter 3.

Overall, the current proposed model is an alternative, new perspective on how the octave illusion is understood. Furthermore, a variant of this model may also potentially describe other illusory stimuli like the scale illusion (Deutsch, 1975), which is also an illusion based on binaural rivalry.

Chapter 7 GENERAL DISCUSSION

Summary of findings

The results from all the experiments (Chapters 2 - 5) suggest a modulation of the auditory percept by directed attention, seen both in the purely sequential ABA_ tone stimulus used in chapter 2 as well as the manipulation of the illusory percepts involving alternating as well as synchronous sound segregation investigated in chapters 3, 4 and 5. Through the experiments of Chapters 3 - 5, we exploit the stimulus based on the octave illusion stimulus to study stream segregation using EEG and psychophysics. This stimulus can be used to understand nuances of sequential and concurrent sound segregation as well as study the effect of attention in complex stream segregation. The results of Chapters 3 - 5 seem to indicate a new aspect of concurrent sound segregation that has previously not been studied. The data opens up new aspects related to context effects, attention, and the role of temporal coherence, which could lead to interesting further research.

7.1 Effect of attention in stream segregation

It has been suggested that auditory attention enables a listener to quickly and accurately direct their 'acoustic searchlight' towards signals of interest in an auditory scene (Fritz et al., 2007). Attention also has been investigated as an on-going, dynamic process leading to selection of information (Carrasco, 2011) based on the confines of an individual's capacity to process sensory information as well as external factors such as context effects and saliency (Corbetta et al., 2008). Knudsen (2007) has proposed that in order to efficiently use attention resources in complex auditory scenarios, listeners must be able to i) select as well as parse out objects of interest based on certain acoustic features such as pitch, localisation, etc. and ii) be adaptable to either maintaining or switching attention between auditory objects in an on-going auditory scene (Shinn-Cunningham, 2008). Many studies have investigated the modulatory effects of attention in selective attention tasks in humans using neuroimaging techniques (see review by Lee et al. (2014)) and have found several neural correlates and evoked responses that are strongly modulated by attention in EEG, MEG and fMRI measures (Hillyard et al., 1973; Woldorff et al., 1993; Petkov et al., 2004).

The data from all the experimental chapters in this thesis have indicated that directed attention could manipulate the percept as well as degree of sound segregation. The first experiment described in chapter 2 revisits the effect of directed attention on sequential stream segregation for a frequency separation within the bistable perceptual region using a classical ABA_ tone triplet stimulus (previously studied by Gutschalk et al., 2005; Snyder et al., 2006, Thompson et al., 2010). Despite the lack of an objective deviant detection task, a significant albeit small effect of attention was noted. Furthermore, the effect of directed attention on the build-up of segregation was distinctly seen in the subjective responses.

Chapters 3, 4 and 5 study the effect of directed attention on a new stimulus similar to that used to elicit the octave illusion (Deutsch, 1974). The role of attention in the perception of the octave illusion has not been studied in detail as the majority of the research focuses on the spontaneous percept elicited by the illusory percept (Deutsch, 1981; Lamminmäki and Hari, 2000; Brancucci et al., 2009, 2011). Furthermore, no task related objective measures have previously been used with this illusion. The relationship between this illusion and streaming has been briefly speculated upon previously (Lamminmäki and Hari, 2000; Micheyl et al., 2010; Schwartz et al., 2012); however, there have not been any dedicated studies to look at the effect of directed attention and the role of streaming in the octave illusion. The data in this thesis suggests that the octave illusion can be manipulated successfully by directing attention using a simple tonal priming sequence. This key finding makes this stimulus more accessible for use as a stream segregation stimulus as the percept is not unchanging for a given individual as previous studies suggest. Chapter 3 also demonstrated an interesting finding where the brain activation at the rate of the illusory percept was only seen for the subset of trials where the objective deviant detection task was correctly carried out.

7.2 New multistable stimulus to study stream segregation

Schwartz et al. (2012) stress on the need for extending the range of stimuli used to study multistability in the auditory domain. They report that not many stimuli used to study auditory multistability allow the studying of binaural rivalry, which would be equivalent to binocular rivalry in the visual domain. Their overview also lists Deutsch’s octave illusions and scale illusions as stimuli that could probably be used to look into auditory multistable perception involving inter-aural grouping.

Several recent studies have started looking at selective attention in dichotic streaming paradigms (Bharadwaj et al., 2014; Choi et al., 2013; Ding and Simon, 2012). However, truly coherent, dichotic streaming stimuli have been typically studied with single tone streams in each ear to establish the extent of perceptual segregation in sounds that are synchronous (Micheyl et al., 2010; 2013a; 2013b).

The octave illusion stimulus presents a simple yet unique setup of tonal stimuli where in order to parse a single tone out of the tonal complex, the listener would need to carry out both within-ear as well as across ear segregation. This stimulus provides a simple, yet previously unexplored opportunity to study auditory stream segregation for alternating as well as synchronous sounds. Furthermore, the stimulus itself is acoustically simple as it only consists of two pure tones of distinctly different frequencies, which in isolation are easily separable.

One might argue the utility of using such a stimulus in understanding realistic auditory environments. However, as noted in section 1.2.5, studying illusions and confusions related to any sensory domain is essential as the resultant illusory or anomalistic percepts help to isolate and clarify the fundamental processes that conventionally lead to accuracy of perception and appropriate interpretation of ambiguous sounds (Warren & Warren, 1970).

7.3 The octave Illusion and the dual mechanism model

The octave illusion stimulus suggested by Deutsch (1974) results in a relatively unexpected percept from the point of view of typical streaming theories. The model proposed by Deutsch (1981), which builds on the theory of separate ‘what’ and ‘where’ pathways (Rauschecker & Scott, 2009) is contrary to a response expected for sounds that are concurrently presented (Elhilali et al., 2009). It has been shown that sounds that are temporally coherent are difficult to stream apart and are usually perceived as a single stream (Elhilali et al., 2009; Shamma et al., 2011; 2013; Micheyl et al., 2010; 2013). Keeping this in mind, the absolute suppression of the pitch percept of the concurrent tones in one of the ears (as suggested by the dual mechanism model) seems unlikely.

The data from chapters 4 and 5 suggest an alternative understanding of the octave illusion in that the synchronous tone pairs were responsible for the illusory percept. Furthermore, in chapter 6, a new model is proposed that explains the illusion in light of competition across binaural cells.

7.4 Role of temporal coherence in the octave illusion

Recent models of segregation and binding have explored the role of temporal structure in binding in vision (Treisman, 1999; Blake & Lee, 2005) and audition (Elhilali et al., 2009; Shamma et al., 2011; 2013). The Gestalt principle of common fate had been employed in the perceptual analysis of acoustic scenes (van Noorden, 1975; Bregman, 1990). Sounds that start and stop together are said to share a common fate and can be attributed to the same acoustic source. A source of sound is associated with several spectro-temporal properties such as pitch, and intensity that co-vary together in time, and this temporal feature can be exploited for segregation. Most models of stream segregation attribute a predominant role in segregation to spectral properties (Fishman et al., 2001; Micheyl et al., 2005; 2007; Bee and Klump, 2004; 2005).

Elhilali and colleagues (2009) demonstrated the shortcomings of the population-separation model by showing that alternating and synchronous patterns of tones produce the same response profiles in the auditory cortex although they have different perceptual signatures: the alternating sequence of tones is perceived as two streams whilst the synchronous sequence of tones is perceived as a single stream. These findings suggest the importance of temporal structure in auditory segregation: sound elements with high temporal coherence may be grouped as one stream whilst elements with low coherence are perceived as separate streams (Shamma et al., 2011). This model of segregation, known as the 'temporal coherence' model, stresses the importance of temporal features in addition to spectral features in determining the perceptual representation of sound scenes.

The octave illusion, however, provides a specific case where the temporal coherence between two concurrent tones breaks down. However, this is only limited to the unique condition where the neighbouring coherent tone pairs are exactly spectrally inverse, as in the case of the octave illusion.

Chapter 6 proposes that this occurs due to competition between the binaural cells and it has also been illustrated that this competition across binaural cells would not occur when the frequencies of the tones are not according to the octave illusion stimulus setup (see Chapter 6). This provides us with a new understanding of a specific condition in which the temporal coherence between concurrent tones is overruled.

7.5 Future directions for research

A key result of the studies described in this thesis was the understanding of a new stimulus that could be used to study sound segregation for alternating as well as synchronous sounds. It has been seen that this stimulus is malleable and the percept of most participants can be manipulated by simple priming sequences. However, these studies have only scratched the surface of all the experimental manipulations possible with this stimulus arrangement.

A primary avenue to further explore would be that of the effects of context on this percept. It has been seen that this illusory percept can only be elicited within certain contextual parameters (see Chapter 4). It would be interesting to do a systematic study on what amount of jitter in these contextual parameters causes the percept to break down. For example, it will be interesting to investigate how the illusion works for tone pairs that are not exactly inverse in frequency but are slightly deviant from the exact low and high frequencies. It would also be interesting to see if prolonged exposure to slightly mismatched frequency pairs leads to the illusory percept over time.

Another possible investigation concerns the representation of the unattended tone pairs that form the background. It has been noted that they do not contribute to the illusory percept per se but are essential to the perception of the illusion.

Lastly, one could investigate the crossover of the findings from this illusory stimulus with other illusory stimuli like Deutsch's Scale Illusion (Deutsch, 1975). The scale illusion also incorporates inter-aural grouping across bilaterally concurrent tones with an additional factor of grouping according to pitch similarities across ears. Understanding this illusion from a streaming perspective would lead to a potentially new stimulus that could be used to understand complex stream segregation scenarios.

Chapter 8 REFERENCES

Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research*, *229*(1–2), 225–236. http://doi.org/10.1016/j.heares.2007.01.011

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.

Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience*, *15*(7), 1063–1073. http://doi.org/10.1162/089892903770007443

Alain, C., Schuler, B. M., & McDonald, K. L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *The Journal of the Acoustical Society of America*, *111*(2), 990–995.

Alain, C., & Woods, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology*, *34*(5), 534–546. http://doi.org/10.1111/j.1469-8986.1997.tb01740.x

Anstis, S. M., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(3), 257–271. http://doi.org/10.1037/0096-1523.11.3.257

Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *The Journal of the Acoustical Society of America*, *88*(2), 680–697. http://doi.org/10.1121/1.399772

Assmann, P. F., & Summerfield, Q. (1994). The contribution of waveform interactions to the perception of concurrent vowels. *The Journal of the Acoustical Society of America*, *95*(1), 471–484. http://doi.org/10.1121/1.408342

Beauvois, M. W., & Meddis, R. (1991). A computer model of auditory stream segregation. *The Quarterly Journal of Experimental Psychology Section A*, *43*(3), 517–541. http://doi.org/10.1080/14640749108400985

Beauvois, M. W., & Meddis, R. (1996). Computer simulation of auditory stream segregation in alternating‐tone sequences. *The Journal of the Acoustical Society of America*, *99*(4), 2270–2280. http://doi.org/10.1121/1.415414

Bee, M. A., & Klump, G. M. (2004). Primitive Auditory Stream Segregation: A Neurophysiological Study in the Songbird Forebrain. *Journal of Neurophysiology*, *92*(2), 1088–1104. http://doi.org/10.1152/jn.00884.2003

Bee, M. A., & Klump, G. M. (2005). Auditory Stream Segregation in the Songbird Forebrain: Effects of Time Intervals on Responses to Interleaved Tone Sequences. *Brain, Behavior and Evolution*, *66*(3), 197–214. http://doi.org/10.1159/000087854

Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception & Psychophysics*, *64*(5), 844–854. http://doi.org/10.3758/BF03194750

Bharadwaj, H. M., Lee, A. K. C., & Shinn-Cunningham, B. G. (2014). Measuring auditory selective attention using frequency tagging. *Frontiers in Integrative Neuroscience*, *8*, 6. http://doi.org/10.3389/fnint.2014.00006

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.-E., Giard, M.-H., & Bertrand, O. (2007). Effects of Selective Attention on the Electrophysiological Representation of Concurrent Sounds in the Human Auditory Cortex. *The Journal of Neuroscience*, *27*(35), 9252–9261. http://doi.org/10.1523/JNEUROSCI.1402-07.2007

Billig, A. J., Davis, M. H., Deeks, J. M., Monstrey, J., & Carlyon, R. P. (2013). Lexical Influences on Auditory Streaming. *Current Biology*, *23*(16), 1585–1589. http://doi.org/10.1016/j.cub.2013.06.042

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, *14*(10), 693–707. http://doi.org/10.1038/nrn3565

Blake, R., & Lee, S.-H. (2005). The Role of Temporal Structure in Human Vision. *Behavioral and Cognitive Neuroscience Reviews*, *4*(1), 21–42. http://doi.org/10.1177/1534582305276839

Boly, M., Garrido, M. I., Gosseries, O., Bruno, M.-A., Boveroux, P., Schnakers, C., … Friston, K. (2011). Preserved Feedforward But Impaired Top-Down Processes in the Vegetative State. *Science*, *332*(6031), 858–862. http://doi.org/10.1126/science.1202043

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. http://doi.org/10.1163/156856897X00357

Brancucci, A., Lugli, V., Perrucci, M. G., Del Gratta, C., & Tommasi, L. (2014). A frontal but not parietal neural correlate of auditory consciousness. *Brain Structure & Function*. http://doi.org/10.1007/s00429-014-0918-2

Brancucci, A., Lugli, V., Santucci, A., & Tommasi, L. (2011). Ear and pitch segregation in Deutsch's octave illusion persist following switch from stimulus alternation to repetition. *The Journal of the Acoustical Society of America*, *130*(4), 2179–2185. http://doi.org/10.1121/1.3631665

Brancucci, A., Padulo, C., & Tommasi, L. (2009). “Octave illusion” or “Deutsch’s illusion”? *Psychological Research PRPF*, *73*(3), 303–307. http://doi.org/10.1007/s00426-008-0153-7

Brancucci, A., Prete, G., Meraglia, E., di Domenico, A., Lugli, V., Penolazzi, B., & Tommasi, L. (2012). Asymmetric Cortical Adaptation Effects during Alternating Auditory Stimulation. *PLoS ONE*, *7*(3), e34367. http://doi.org/10.1371/journal.pone.0034367

Brännström, K. J., & Nilsson, P. (2011). Octave illusion elicited by overlapping narrowband noises. *The Journal of the Acoustical Society of America*, *129*(5), 3213–3220. http://doi.org/10.1121/1.3571425

Bregman, A. S. (1978a). Auditory streaming: Competition among alternative organizations. *Perception & Psychophysics*, *23*(5), 391–398. http://doi.org/10.3758/BF03204141

Bregman, A. S. (1978b). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(3), 380–387. http://doi.org/10.1037/0096-1523.4.3.380

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

Bregman, A. S., Ahad, P. A., Crum, P. A. C., & O’Reilly, J. (2000). Effects of time intervals and tone durations on auditory stream segregation. *Perception & Psychophysics*, *62*(3), 626–636. http://doi.org/10.3758/BF03212114

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*(2), 244–249. http://doi.org/10.1037/h0031163

Bregman, A. S., & Rudnicky, A. I. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 263–267. http://doi.org/10.1037/0096-1523.1.3.263

Bregman, A. S., & Steiger, H. (1980). Auditory streaming and vertical localization: Interdependence of “what” and “where” decisions in audition. *Perception & Psychophysics*, *28*(6), 539–546. http://doi.org/10.3758/BF03198822

Brochard, R., Drake, C., Botte, M.-C., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1742–1759. http://doi.org/10.1037/0096-1523.25.6.1742

Carlyon, R. P. (2004a). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*(10), 465–471. http://doi.org/10.1016/j.tics.2004.08.008

Carlyon, R. P. (2004b). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*(10), 465–471. http://doi.org/10.1016/j.tics.2004.08.008

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. http://doi.org/10.1037/0096-1523.27.1.115

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. http://doi.org/10.1016/j.visres.2011.04.012

Chait, M., Poeppel, D., & Simon, J. Z. (2008). Auditory temporal edge detection in human auditory cortex. *Brain Research*, *1213*, 78–90. http://doi.org/10.1016/j.brainres.2008.03.050

Chalikia, M. H., & Bregman, A. S. (1989). The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation. *Perception & Psychophysics*, *46*(5), 487–496. http://doi.org/10.3758/BF03210865

Chalikia, M. H., & Bregman, A. S. (1993). The perceptual segregation of simultaneous vowels with harmonic, shifted, or random components. *Perception & Psychophysics*, *53*(2), 125–133. http://doi.org/10.3758/BF03211722

Chambers, C. D., Mattingley, J. B., & Moss, S. A. (2002). The octave illusion revisited: Suppression or fusion between ears? *Journal of Experimental Psychology: Human Perception and Performance*, *28*(6), 1288–1302. http://doi.org/10.1037/0096-1523.28.6.1288

Chambers, C. D., Mattingley, J. B., & Moss, S. A. (2004). Reconsidering evidence for the suppression model of the octave illusion. *Psychonomic Bulletin & Review*, *11*(4), 642–666. http://doi.org/10.3758/BF03196617

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. http://doi.org/10.1121/1.1907229

Choi, I., Wang, L., Bharadwaj, H., & Shinn-Cunningham, B. (2014). Individual differences in attentional modulation of cortical responses correlate with selective attention performance. *Hearing Research*, *314*, 10–19. http://doi.org/10.1016/j.heares.2014.04.008

Christensen, I. P., & Gregory, A. H. (1977). Further study of an auditory illusion. *Nature*, *268*(5621), 630–631. http://doi.org/10.1038/268630a0

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306–324. http://doi.org/10.1016/j.neuron.2008.04.017

Cusack, R., Decks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of Location, Frequency Region, and Time Course of Selective Attention on Auditory Scene Analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 643–656. http://doi.org/10.1037/0096-1523.30.4.643

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, *62*(5), 1112–1120.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In *The handbook of perception and cognition: Hearing. B.C.J. Moore* (p. 6: 387–424.). London, Academic Press.

Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 617–629. http://doi.org/10.1037/0096-1523.25.3.617

Darwin, C. J., Hukin, R. W., & Al-Khatib, B. Y. (1995). Grouping in pitch perception: Evidence for sequential constraints. *The Journal of the Acoustical Society of America*, *98*(2), 880–885. http://doi.org/10.1121/1.413513

Deike, S., Heil, P., Böckmann-Barthel, M., & Brechmann, A. (2012). The Build-up of Auditory Stream Segregation: A Different Perspective. *Frontiers in Psychology*, *3*. http://doi.org/10.3389/fpsyg.2012.00461

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

Denham, S. L., Gyimesi, K., Stefanics, G., & Winkler, I. (2010). Stability of Perceptual Organisation in Auditory Streaming. In E. A. Lopez-Poveda, A. R. Palmer, & R. Meddis (Eds.), *The Neurophysiological Bases of Auditory Perception* (pp. 477–487). Springer New York.

Denham, S. L., & Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *Journal of Physiology-Paris*, *100*(1–3), 154–170. http://doi.org/10.1016/j.jphysparis.2006.09.012

Deutsch, D. (1974). An auditory illusion. *The Journal of the Acoustical Society of America*, *55*(S1), S18–S19. http://doi.org/10.1121/1.1919587

Deutsch, D. (1974). An auditory illusion. *Nature*, *251*(5473), 307–309.

Deutsch, D. (1975). Two-channel listening to musical scales. *The Journal of the Acoustical Society of America*, *57*(5), 1156–1160.

Deutsch, D. (1978). Lateralization by frequency for repeating sequences of dichotic 400- and 800-Hz tones. *The Journal of the Acoustical Society of America*, *63*(1), 184–186. http://doi.org/10.1121/1.381710

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, *28*(5), 381–389. http://doi.org/10.3758/BF03204881

Deutsch, D. (1981). The octave illusion and auditory perceptual integration. *Hearing Research and Theory*, *1*, 99–142.

Deutsch, D. (1983). The generation of two isochronous sequences in parallel. *Perception & Psychophysics*, *34*(4), 331–337. http://doi.org/10.3758/BF03203045

Deutsch, D. (1987). The tritone paradox: Effects of spectral variables. *Perception & Psychophysics*, *41*(6), 563–575. http://doi.org/10.3758/BF03210490

Deutsch, D. (1988). Lateralization and sequential relationships in the octave illusion. *The Journal of the Acoustical Society of America*, *83*(1), 365–369. http://doi.org/10.1121/1.396249

Deutsch, D. (1995). *Musical illusions and paradoxes*. Philomel.

Deutsch, D. (2004). The Octave Illusion Revisited Again. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 355–364. http://doi.org/10.1037/0096-1523.30.2.355

Deutsch, D., & Gregory, A. H. (1978). Deutsch's octave illusion. *Nature*, *274*(5672), 721–721. http://doi.org/10.1038/274721b0

Deutsch, D., & Roll, P. L. (1976). Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance*, 2(1), 23–29. http://doi.org/10.1037/0096-1523.2.1.23

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, *109*(29), 11854–11859. http://doi.org/10.1073/pnas.1205381109

Dyson, B. J., & Alain, C. (2004). Representation of concurrent acoustic objects in primary auditory cortex. *The Journal of the Acoustical Society of America*, *115*(1), 280–288.

Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC Press.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., & Shamma, S. A. (2009). Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. *Neuron*, *61*(2), 317–329. http://doi.org/10.1016/j.neuron.2008.12.005

Elhilali, M., & Shamma, S. A. (2008). A cocktail party with a cortical twist: How cortical mechanisms contribute to sound segregation. *The Journal of the Acoustical Society of America*, *124*(6), 3751–3771. http://doi.org/10.1121/1.3001672

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research*, *151*(1-2), 167–187. http://doi.org/10.1016/S0378-5955(00)00224-0

Fishman, Y. I., & Steinschneider, M. (2010). Neural correlates of auditory scene analysis based on inharmonicity in monkey primary auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(37), 12480–12494. http://doi.org/10.1523/JNEUROSCI.1780-10.2010

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention — focusing the searchlight on sound. *Current Opinion in Neurobiology*, *17*(4), 437–455. http://doi.org/10.1016/j.conb.2007.07.011

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*, *78*(4), 2643–2647.

Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous Cortical Rhythms Determine Cerebral Specialization for Speech Perception and Production. *Neuron*, *56*(6), 1127–1134. http://doi.org/10.1016/j.neuron.2007.09.038

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. http://doi.org/10.1038/nn.3063

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*(11), 887–892. http://doi.org/10.1038/nrn1538

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal of the Acoustical Society of America*, *111*(3), 1340–1348. http://doi.org/10.1121/1.1452740

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2000). Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *The Journal of the Acoustical Society of America*, *108*(1), 263–271. http://doi.org/10.1121/1.429462

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., & Oxenham, A. J. (2005). Neuromagnetic Correlates of Streaming in Human Auditory Cortex. *The Journal of Neuroscience*, *25*(22), 5382–5388. http://doi.org/10.1523/JNEUROSCI.0347-05.2005

Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C., & Melcher, J. R. (2007). Human Cortical Activity during Streaming without Spectral Cues Suggests a General Neural Substrate for Auditory Stream Segregation. *The Journal of Neuroscience*, *27*(48), 13074–13081. http://doi.org/10.1523/JNEUROSCI.2299-07.2007

Hari, R., Hämäläinen, M., & Joutsiniemi, S.-L. (1989). Neuromagnetic steady - state responses to auditory stimuli. *The Journal of the Acoustical Society of America*, *86*(3), 1033–1039. http://doi.org/10.1121/1.398093

Hartmann, W. M., & Johnson, D. (1991). Stream Segregation and Peripheral Channeling. *Music Perception: An Interdisciplinary Journal*, *9*(2), 155–183. http://doi.org/10.2307/40285527

Haywood, N. R., & Roberts, B. (2013). Build-up of auditory stream segregation induced by tone sequences of constant or alternating frequency and the resetting effects of single deviants. *Proceedings of Meetings on Acoustics*, *19*(1), 050080. http://doi.org/10.1121/1.4799220

Helmholtz, H. von. (1925). Physiological optics. *Optical Society of America*, *3*, 318.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science (New York, N.Y.)*, *182*(4108), 177–180.

Hupé, J.-M., & Pressnitzer, D. (2012). The initial phase of auditory and visual scene analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 942–953. http://doi.org/10.1098/rstb.2011.0368

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. http://doi.org/10.1037/0096-1523.21.4.751

Jones, D., Alford, D., Bridges, A., Tremblay, S., & Macken, B. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 464–473. http://doi.org/10.1037/0278-7393.25.2.464

Joris, P., & Yin, T. C. T. (2007). A matter of time: internal delays in binaural processing. *Trends in Neurosciences*, *30*(2), 70–78. http://doi.org/10.1016/j.tins.2006.12.004

Kaneko, K., Fujiki, N., & Hari, R. (2003). Binaural interaction in the human auditory cortex revealed by neuromagnetic frequency tagging: no effect of stimulus intensity. *Hearing Research*, *183*(1–2), 1–6. http://doi.org/10.1016/S0378-5955(03)00186-2

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional Gain Control of Ongoing Cortical Speech Representations in a “Cocktail Party.” *The Journal of Neuroscience*, *30*(2), 620–628. http://doi.org/10.1523/JNEUROSCI.3631-09.2010

Knudsen, E. I. (2007). Fundamental Components of Attention. *Annual Review of Neuroscience*, *30*(1), 57–78. http://doi.org/10.1146/annurev.neuro.30.051606.094256

Koffka, K. (1935). *Principles of gestalt psychology.* New York: Harcourt, Brace & World.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*(1–2), 97–126. http://doi.org/10.1016/S0010-0277(00)00155-4

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. *Science*, *320*(5872), 110–113. http://doi.org/10.1126/science.1154735

Lamminmäki, S., & Hari, R. (2000). Auditory cortex activation associated with octave illusion. *Neuroreport*, *11*(7), 1469–1472.

Lamminmäki, S., Mandel, A., Parkkonen, L., & Hari, R. (2012). Binaural interaction and the octave illusion. *The Journal of the Acoustical Society of America*, *132*(3), 1747–1753. http://doi.org/10.1121/1.4740474

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82. http://doi.org/10.1016/j.tics.2004.12.004

Lee, A. K. C., Larson, E., Maddox, R. K., & Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hearing Research*, *307*, 111–120. http://doi.org/10.1016/j.heares.2013.06.010

Lee, A. K. C., Rajaram, S., Xia, J., Bharadwaj, H., Larson, E., Hamalainen, M. S., & Shinn-Cunningham, B. G. (2013). Auditory Selective Attention Reveals Preparatory Activity in Different Cortical Regions for Selection Based on Source Location and Source Pitch. *Frontiers in Neuroscience*, *6*. http://doi.org/10.3389/fnins.2012.00190

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences*, *3*(7), 254–264.

Linke, A. C., Vicente-Grabovetsky, A., & Cusack, R. (2011). Stimulus-specific suppression preserves information in auditory short-term memory. *Proceedings of the National Academy of Sciences*, *108*(31), 12961–12966. http://doi.org/10.1073/pnas.1102118108

Lipp, R., Kitterick, P., Summerfield, Q., Bailey, P. J., & Paul-Jordanov, I. (2010). Concurrent sound segregation based on inharmonicity and onset asynchrony. *Neuropsychologia*, *48*(5), 1417–1425. http://doi.org/10.1016/j.neuropsychologia.2010.01.009

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010. http://doi.org/10.1016/j.neuron.2007.06.004

MacDougall-Shackleton, S. A., Hulse, S. H., Gentner, T. Q., & White, W. (1998). Auditory scene analysis by European starlings (Sturnus vulgaris): Perceptual segregation of tone sequences. *The Journal of the Acoustical Society of America*, *103*(6), 3581–3587. http://doi.org/10.1121/1.423063

Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P., & Jones, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 43–51. http://doi.org/10.1037/0096-1523.29.1.43

McAdams, S., & Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *The Journal of the Acoustical Society of America*, *102*(5 Pt 1), 2945–2953.

McCabe, S. L., & Denham, M. J. (1997). A model of auditory streaming. *The Journal of the Acoustical Society of America*, *101*(3), 1611–1621. http://doi.org/10.1121/1.418176

McClurkin, R. H., & Iii, J. W. H. (1981). Pitch and timbre in a two‐tone dichotic auditory illusion. *The Journal of the Acoustical Society of America*, *69*(2), 592–594. http://doi.org/10.1121/1.385376

McDermott, J. H. (2009). The cocktail party problem. *Current Biology: CB*, *19*(22), R1024–1027. http://doi.org/10.1016/j.cub.2009.09.005

McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience*, *16*(4), 493–498. http://doi.org/10.1038/nn.3347

McDermott, J. H., & Simoncelli, E. P. (2011). Sound Texture Perception via Statistics of the Auditory Periphery: Evidence from Sound Synthesis. *Neuron*, *71*(5), 926–940. http://doi.org/10.1016/j.neuron.2011.06.032

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*(7397), 233–236. http://doi.org/10.1038/nature11020

Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., … Courtenay Wilson, E. (2007). The role of auditory cortex in the formation of auditory streams. *Hearing Research*, *229*(1–2), 116–131. http://doi.org/10.1016/j.heares.2007.01.007

Micheyl, C., Hanson, C., Demany, L., Shamma, S., & Oxenham, A. J. (2013). Auditory stream segregation for alternating and synchronous tones. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(6), 1568–1580. http://doi.org/10.1037/a0032241

Micheyl, C., Kreft, H., Shamma, S., & Oxenham, A. J. (2013). Temporal coherence versus harmonicity in auditory stream formation. *The Journal of the Acoustical Society of America*, *133*(3), EL188–EL194. http://doi.org/10.1121/1.4789866

Micheyl, C., & Oxenham, A. J. (2010). Objective and Subjective Psychophysical Measures of Auditory Stream Integration and Segregation. *Journal of the Association for Research in Otolaryngology*, *11*(4), 709–724. http://doi.org/10.1007/s10162-010-0227-2

Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual Organization of Tone Sequences in the Auditory Cortex of Awake Macaques. *Neuron*, *48*(1), 139–148. http://doi.org/10.1016/j.neuron.2005.08.039

Middlebrooks, J. C., Dykes, R. W., & Merzenich, M. M. (1980). Binaural response-specific bands in primary auditory cortex (AI) of the cat: Topographical organization orthogonal to isofrequency contours. *Brain Research*, *181*(1), 31–48. http://doi.org/10.1016/0006-8993(80)91257-3

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin*, *44*(2), 105–129. http://doi.org/10.1037/h0055960

Miller, G. A., & Heise, G. A. (1950). The Trill Threshold. *The Journal of the Acoustical Society of America*, *22*(5), 637–638. http://doi.org/10.1121/1.1906663

Mill, R. W., Bőhm, T. M., Bendixen, A., Winkler, I., & Denham, S. L. (2013). Modelling the Emergence and Dynamics of Perceptual Organisation in Auditory Streaming. *PLoS Comput Biol*, *9*(3), e1002925. http://doi.org/10.1371/journal.pcbi.1002925

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *The Journal of the Acoustical Society of America*, *80*(2), 479–483. http://doi.org/10.1121/1.394043

Moore, B. C. J., & Gockel, H. (2002). Factors Influencing Sequential Stream Segregation. *Acta Acustica United with Acustica*, *88*(3), 320–333.

Moore, B. C. J., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 919–931. http://doi.org/10.1098/rstb.2011.0355

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. http://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E. N. (1988). Frequency and location specificify of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology*, *69*(6), 523–531. http://doi.org/10.1016/0013-4694(88)90164-2

Necker, L. A. (1832). LXI. Observations on some remarkable optical phænomena seen in Switzerland; and on an optical phænomenon which occurs on viewing a figure of a crystal or geometrical solid. *Philosophical Magazine Series 3*, *1*(5), 329–337. http://doi.org/10.1080/14786443208647909

Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Current Opinion in Neurobiology*, *14*(4), 474–480. http://doi.org/10.1016/j.conb.2004.06.005

Overath, T., Kumar, S., Kriegstein, K. von, & Griffiths, T. D. (2008). Encoding of Spectral Correlation over Time in Auditory Cortex. *The Journal of Neuroscience*, *28*(49), 13268–13273. http://doi.org/10.1523/JNEUROSCI.4596-08.2008

Overath, T., Kumar, S., Stewart, L., Kriegstein, K. von, Cusack, R., Rees, A., & Griffiths, T. D. (2010). Cortical Mechanisms for the Segregation and Representation of Acoustic Textures. *The Journal of Neuroscience*, *30*(6), 2070–2076. http://doi.org/10.1523/JNEUROSCI.5378-09.2010

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. http://doi.org/10.1163/156856897X00366

Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience*, *7*(6), 658–663. http://doi.org/10.1038/nn1256

Picton, T. W., Woods, D. L., & Proulx, G. B. (1978). Human auditory sustained potentials. I. The nature of the response. *Electroencephalography and Clinical Neurophysiology*, *45*(2), 186–197. http://doi.org/10.1016/0013-4694(78)90003-2

Plomp, R., & Mimpen, A. M. (1968). The Ear as a Frequency Analyzer. II. *The Journal of the Acoustical Society of America*, *43*(4), 764–767. http://doi.org/10.1121/1.1910894

Poljak, S. (1927). An experimental study of the association callosal, and projection fibers of the cerebral cortex of the cat. *The Journal of Comparative Neurology*, *44*(2), 197–258. http://doi.org/10.1002/cne.900440202

Pressnitzer, D., & Hupé, J.-M. (2006). Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization. *Current Biology*, *16*(13), 1351–1357. http://doi.org/10.1016/j.cub.2006.05.054

Pressnitzer, D., Suied, C., & Shamma, S. A. (2011). Auditory Scene Analysis: The Sweet Music of Ambiguity. *Frontiers in Human Neuroscience*, *5*. http://doi.org/10.3389/fnhum.2011.00158

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. http://doi.org/10.1038/nn.2331

Razak, K. A. (2011). Systematic Representation of Sound Locations in the Primary Auditory Cortex. *The Journal of Neuroscience*, *31*(39), 13848–13859. http://doi.org/10.1523/JNEUROSCI.1937-11.2011

Ritter, W., Vaughan Jr., H. G., & Costa, L. D. (1968). Orienting and habituation to auditory stimuli: A study of short terms changes in average evoked responses. *Electroencephalography and Clinical Neurophysiology*, *25*(6), 550–556. http://doi.org/10.1016/0013-4694(68)90234-4

Roberts, B., Glasberg, B. R., & J, C. (2008). Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 992–1006. http://doi.org/10.1037/0096-1523.34.4.992

Roberts, B., Glasberg, B. R., & Moore, B. C. J. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *The Journal of the Acoustical Society of America*, *112*(5), 2074–2085. http://doi.org/10.1121/1.1508784

Schneider, G. E. (1969). Two visual systems. *Science*, *163*(3870), 895–902. http://doi.org/10.1126/science.163.3870.895

Schwartz, J.-L., Grimault, N., Hupé, J.-M., Moore, B. C. J., & Pressnitzer, D. (2012). Multistability in perception: binding sensory modalities, an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 896–905. http://doi.org/10.1098/rstb.2011.0254

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114–123. http://doi.org/10.1016/j.tins.2010.11.002

Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology*, *20*(3), 361–366. http://doi.org/10.1016/j.conb.2010.03.009

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. http://doi.org/10.1016/j.tics.2008.02.003

Shinn-Cunningham, B. G., Lee, A. K. C., & Oxenham, A. J. (2007). A sound element gets lost in perceptual competition. *Proceedings of the National Academy of Sciences*, *104*(29), 12223–12227. http://doi.org/10.1073/pnas.0704641104

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of Attention on Neuroelectric Correlates of Auditory Stream Segregation. *Journal of Cognitive Neuroscience*, *18*(1), 1–13. http://doi.org/10.1162/089892906775250021

Stainsby, T. H., Füllgrabe, C., Flanagan, H. J., Waldman, S. K., & Moore, B. C. J. (2011). Sequential streaming due to manipulation of interaural time differencesa). *The Journal of the Acoustical Society of America*, *130*(2), 904–914. http://doi.org/10.1121/1.3605540

Sussman, E., Ritter, W., & Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, *36*(01), 22–34. http://doi.org/null

Sussman, E. S. (2007). A New View on the MMN and Attention Debate: The Role of Context in Processing Auditory Events. *Journal of Psychophysiology*, *21*(3), 164–175. http://doi.org/10.1027/0269-8803.21.34.164

Sussman, E. S., Bregman, A. S., Wang, W. J., & Khan, F. J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 93–110. http://doi.org/10.3758/CABN.5.1.93

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

Sussman, E., & Winkler, I. (2001). Dynamic sensory updating in the auditory system. *Cognitive Brain Research*, *12*(3), 431–439. http://doi.org/10.1016/S0926-6410(01)00067-2

Sussman, E., Winkler, I., & Wang, W. (2003). MMN and attention: Competition for deviance detection. *Psychophysiology*, *40*(3), 430–435. http://doi.org/10.1111/1469-8986.00045

Szalárdy, O., Bőhm, T. M., Bendixen, A., & Winkler, I. (2013). Event-related potential correlates of sound organization: Early sensory and late cognitive effects. *Biological Psychology*, *93*(1), 97–104. http://doi.org/10.1016/j.biopsycho.2013.01.015

Thompson, S. K., Carlyon, R. P., & Cusack, R. (2011). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1253–1262. http://doi.org/10.1037/a0021925

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences*, *10*(11), 502–511. http://doi.org/10.1016/j.tics.2006.09.003

Treisman, A. (1999). Solutions to the Binding Problem: Progress through Controversy and Convergence. *Neuron*, *24*(1), 105–125. http://doi.org/10.1016/S0896-6273(00)80826-0 van Noorden. (1975). *Temporal coherence in the perception of tone sequences*. Technische Hogeschool Eindhoven. Retrieved from http://repository.tue.nl/152538

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. *The Journal of the Acoustical Society of America*, *106*(2), 938–945. http://doi.org/10.1121/1.427140

Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *The Journal of the Acoustical Society of America*, *105*(1), 339–346. http://doi.org/10.1121/1.424503

Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). A Precedence Effect in Sound Localization. *The Journal of the Acoustical Society of America*, *21*(4), 468–468. http://doi.org/10.1121/1.1917119

Warren, R. M., & Warren, R. P. (1970). *Auditory illusions and confusions*. WH Freeman.

Winkler, I., Denham, S., Mill, R., Bőhm, T. M., & Bendixen, A. (2012). Multistability in auditory stream segregation: a predictive coding view. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 1001–1012. http://doi.org/10.1098/rstb.2011.0359

Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., … Sussman, E. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences*, *100*(20), 11812–11815. http://doi.org/10.1073/pnas.2031891100

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, *90*(18), 8722–8726.

Zeki, S. (2004). The neurology of ambiguity. *Consciousness and Cognition*, *13*(1), 173–196. http://doi.org/10.1016/j.concog.2003.10.003

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., … Schroeder, C. E. (2013). Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party." *Neuron*, *77*(5), 980–991. http://doi.org/10.1016/j.neuron.2012.12.037

Zurek, P. M. (1987). The Precedence Effect. In W. A. Yost & G. Gourevitch (Eds.), *Directional Hearing* (pp. 85–105). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4612-4738-8_4

Zwicker, T. (1984). Experimente zur dichotischen Oktav-Täuschung. *Acta Acustica United with Acustica*, *55*(2), 128–136.