Apparent Motion and the Tritone Paradox: An EEG Investigation of Novel Bistable Stimuli

_____

A Thesis

Presented to

The Division of Philosophy, Religion, Psychology, & Linguistics

Reed College

_____

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

_____

Gray Davidson

May 2012

Approved for the Division
(Psychology)

_____

Michael Pitts

Acknowledgements

Table of Contents

List of Tables

List of Figures

Abstract

Bistable figures allow researchers to investigate brain responses associated with perception while holding the physical characteristics of a stimulus constant. Historically, most bistable stimuli which have been used in EEG paradigms have been static images and all have been visual. Previous studies have repeatedly found two primary event-related potential (ERP) components associated with changes in perception while viewing bistable figures, the "reversal negativity" (RN) and the "late positive complex" (LPC). The present examination pioneered two novel stimuli, one based on visual apparent motion, and one based on auditory stimuli designed by Diana Deutsch (similar to the Tritone Paradox) both of which involve bistability in the perceived relationship between pairs of successive stimuli. The experimental hypothesis was that both novel stimuli would evoke scalp potentials similar to those seen in prior studies of bistability, while activation in the pre-stimulus interval (the interval between stimulus 1 and stimulus 2 of each pair) was left as an open question. As hypothesized, the visually bistable motion stimulus evoked a reversal negativity (RN), and a late positive component (LPC), and additionally, two positive components were observed in the pre-stimulus interval just prior to perceptual changes which may be related to the intention or anticipation of perceptual reversal. An LPC was also observed in the auditory condition as well as widespread negative components throughout the stimulus presentation interval which may represent an auditory analog of the RN as well as other potentially novel ERP components.

Chapter 1

Introduction

The modern bloom of mathematical methods, technological aids and standardized investigatory practices situates today's psychological investigators, heirs of Aristotle, Descartes, and William James, in a promising time. As never before, theorists are able to contemplate the human brain and mind, supplementing their theories with evidence from a myriad of powerful techniques. The goal of this thesis is to employ a novel combination of established neurophysiological and psychophysical methods to explore the neural basis of conscious perception in both the visual and auditory modalities.

Despite obvious differences between visual and auditory perception, the macroscopic features of the two systems are remarkably similar. Both vision and audition receive information from physical waves in air, each uses an interface with the outside world to filter these physical properties, each employs electrical signaling to transfer the collected information to the brain, and each then has a complex and powerful system to process and sort the information into an intelligible percept which is invaluable for proper functioning and survival.

In the present investigation bistable stimuli and the ERP technique are combined in search of similarities and differences between the neural systems underlying visual and auditory perception. New stimuli have been designed in both modalities, utilizing principles of apparent motion for visual bistability, and the tritone paradox for the auditory bistability. Comparisons will be made in each modality between trials in which the percept switched versus trials in which it remained in its previous state. While many experiments have been performed with static visual bistable stimuli, mapping what may be expected from this paradigm, it is less well known how using a dynamic stimulus will affect the data.

Auditory bistability is unexplored territory. This investigation will shed light on the differences and similarities between auditory and visual perception, and may be able to indicate whether the mechanisms behind bistable perception lie in the sensory systems themselves, or lie in higher processing areas of the brain.

Multistable phenomena are of incredible use in cognitive neuroscience because they overcome one of the most ubiquitous confounds facing researchers. When two distinct images produce two distinct neural representations, it is difficult (and indeed logically impossible) to argue that any particular aspect of the images is responsible

for the difference in the brain – research is always confounded by the presentation of differentiable stimuli. Multistable stimuli solve this problem because although they can be perceived in multiple different ways (generally two), the image itself does not change so any neural differentiation can be attributed to perception, rather than sensation [2, 91].

The distinction between perception and sensation is of paramount importance in the study of human sensory systems, and to this investigation in particular. As Tovee once proposed, "...the nature of our representation may change even when there has been no change in the external stimulus. Perception seems to be less a representation of our environment, but rather an interpretation, and our interpretation may change based on cognitive rather than perceptual factors. The visual stimulus does not have to change for our perception to be transformed [94]."

1.1   Visual Bistability

Vision is by far the best explored of human senses, due in part to its privileged place in the human sensory experience. The visual system maps the incoming light rays from every point in the visual field, using first the simple amplitude, wavelength, and position of the light then seeks pattens within this raw data to form perceptual interpretations of what exists in the outside world [83]. Due to the multilayered nature of this system, and the rules of allocation which it uses to build an object-based interpretation of the world,[1] it can be exploited by a class of images which are called *multi-stable*. These images are constructed such that they can be perceived in multiple, internally consistent ways, without any change in sensory information. Such illusions are not simply artifacts of psychological laboratories, but can occur in important, survival-related situations as in Fig. 1.1 [77].

In terms of motion perception, specifically, the brain is capable of seeing movement when an object disappears from one point and reappears at another in the visual field. At no point was either object seen in motion, although the visual scene has indeed changed, and if the distance is small enough, and if the two events are close enough together in time, the brain will decide that a single object has moved, rather than that a new object has appeared. This fact is employed by the film industry to create a moving picture for the television screen out of a series of distinct frames, and can also be employed by psychologists to test the visual system.

Multistable (and in particular *bi-stable*) images are commonly categorized as visual illusions, and many people are familiar with some of the more famous examples. One of the most common is the Necker cube[2] [68] whose faces can be seen as extending

---

[1]Separate areas of the brain have been identified which are likely responsible for humans' particular acuity when recognizing faces [46] and other extremely familiar sights, such as the look of one's home [93, 38, 40].

[2]For this cube, two opposite faces can equally be seen as the frontal face, while stability in either percept forces the unattended face to assume position as the back face [68]. This and other optical illusions are shown in Fig. 1.3. Another common example is Fig. 1.3 (b), in which a pair of face profiles or a central vase can mutually exclusively be seen as the foreground and background of the
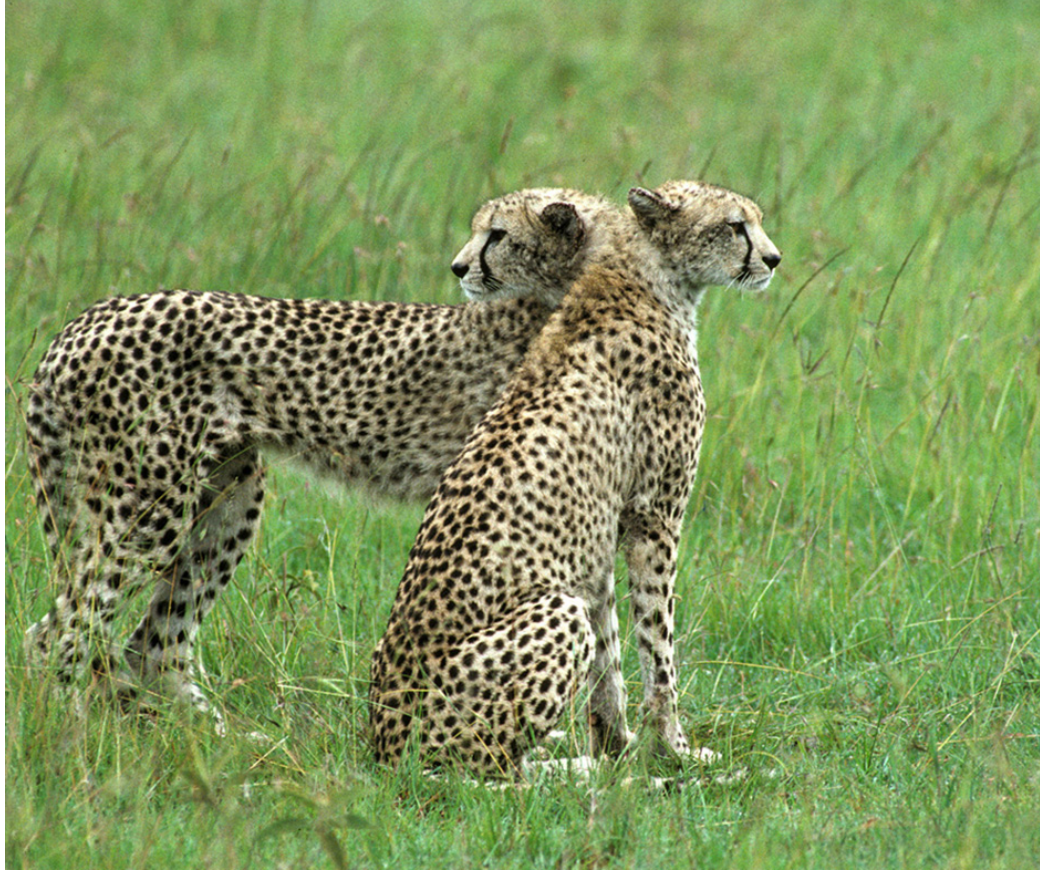
Figure 1.1: A real-world example of visual bistability, originally captured by wildlife photographer Gerry Lemmo, and used in a neuroimaging study of bistability in 2008 [76].

from the page in either of two orientations.

The use of bistable stimuli in perception experiments has become popular in recent years because multiple perceptions can be evoked by identical sensory inputs which allows an exceptional level of experimental control.[3] The Reversal rates, percept durations, and neurological responses to stimuli of this kind help to shed light on the liminal area between sensation and perception and can help to elucidate such important questions as when a certain perception reaches consciousness. When a bistable stimulus is used, differences in recorded electrical or blood oxygen level-dependent (BOLD) activity can be argued to come from the perceptual differences, rather than sensory ones.

---

image [81]. This is an example of exclusive allocation which is a type of visual scene analysis [12] and is responsible for each feature of our environment belonging to only one object or source.

[3]The majority of bistable images are at least partially multistable in that there is usually a third version of the percept which is akin to the raw data input to the sensory system. Those versions of the percept which contain coherent information are more attractive in some sense to the attention of the viewer, and it is sometimes even difficult to break away from these to see the raw form again.

1.1.1   Visual Scene Analysis

In the visual cortex, successive layers perform progressively more complicated functions. First the entire visual field is encoded based on features of light, location, intensity and frequency, then lines and angles are detected and shapes are formed, and finally objects are recognized. Analysis of the visual scene around us is one of the primary tasks of the visual system, dividing it into objects and parsing their physical limits, positions, and whether and how they are moving.

Coherence in terms of color, direction and speed of motion, direction of line, or relationships between lines such as parallelism or perpendicularity are cues to the brain that a given set of features belong to the same object. Circles, for instance, are called 'strong' perceptual forms because they exhibit perceived continuity. Even if a circle's outer edges are incomplete in a given picture, it will not be seen as incomplete, but as continuing on behind the other forms. In other words, the circle has closed perceptually [11]. Sets of objects which move in a coherent manner are likewise often perceived as belonging to the same source.

1.1.2   Static Visual Illusions

Pitting one of its sub-systems against another is a standard way of inducing a visual illusion. A simple example is the checker shadow illusion Fig. 1.2 (a) in which two squares on a checkerboard, a dark square in the sun and a light square in the shade are in fact the exact same color but are perceived differently due to the presence of extra information: the shadow of a cylinder which falls over the light square. Even though the brain has all the necessary *sensory* color information available, a shadow is perceived so a perceptual adjustment is made on how the luminosity information is processed. The illusion is broken in Fig. 1.2 (b).
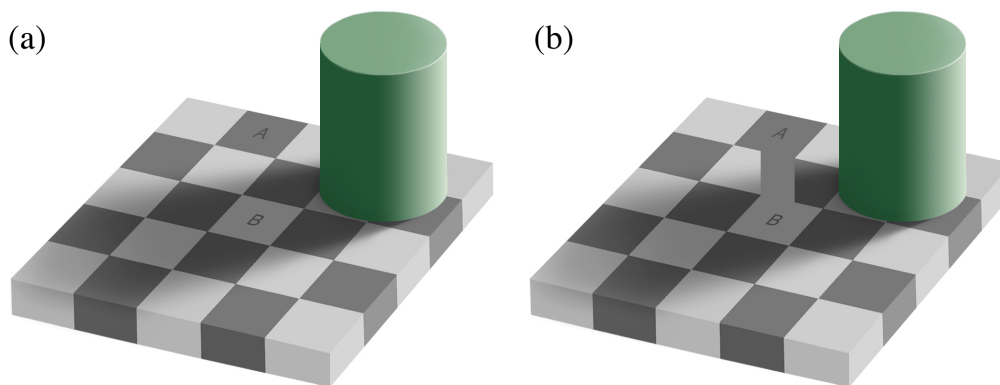
Figure 1.2: The Checkerboard illusion (a) is a simple use of external cues (*i.e.* what we know about light and shade, and what we know about chess boards) to convince our brains that the two indicated squares are different colors when in fact they are identical as can be seen in part (b).

Several classic examples of visual illusions include the Necker cube [68], the Rubin

vase[85], and Schroeder's stairs [7]. These images each employ elements of the visual system in conflict to create an illusion. The last example, for instance, does so by violating standard rules of perspective drawing.

Figure 1.3: Three classic optical illusions, the Necker cube (a), the Rubin vase (b), and the Schroeder stairs (c). All of the information in each of these images can be accounted for by two mutually exclusive interpretations. In the cases of (a) and (c), the interpretation share primary characteristics, but in the case of (b), the entire interpretation of the scene changes.

The examples in Fig. 1.3 (a), (b), and (c) are all categorized as *multistable* because each has multiple, mutually exclusive interpretations, each of which accounts for all the features of the image.

Binocular rivalry is a method of inducing a visual illusion using any two visual stimuli. It has been used in a myriad of studies over the last century, although its initial conception dates much earlier to the 18th or even 16th centuries [100]. The principle is a simple one in which two distinct images are presented, each to a different eye, and the startling result is that only one is available to the consciousness of the viewer at a time.

Binocular rivalry is a limited example of a bistable percept. Two perceptions do occur, but two distinct images are indeed in front of the viewer's eyes, although the paradigm has interesting implications for the study of visual consciousness.[4].

Perhaps the most startling thing about multistable images is that perception alternates inevitably between the possible stabilities. Although it may be odd that the same set of lines on paper can be thought of in two completely different ways, it is certainly stranger that (with continued exposure) it becomes impossible *not* to experience each in turn.

Conflicting sensory and behavioral theories of perceptual alternation have been posited. Initially, it was thought that tiring of feature-related neurons or fatigue

---

[4]It is thought that monocular neurons play a part in the promotion of one stimulus and the suppression of the other [6], but it is likely that only a subset of the monocular neurons corresponding to a specific visual sensation participate in the larger question of whether this sensation becomes perception [59]

Binocular rivalry has been used successfully in both hemodynamic and electrical imaging studies, and although some portion of the mechanics behind the perceptual switching witnessed with this paradigm must be different than that from multistable figures, some aspects of these designs do seem to be similar, such as switching rates between percepts [14, 57].

of other elements of the visual hierarchy were responsible for the perceptual switch
[51], but more recently data have indicated that a behavioral mechanism may be
more logical. This proposition, from Leopold & Logothetis in 1999 was based on
the similarities between the perceptual switching of bistable stimuli, and the results
of neuroimaging studies, analyses of temporal dynamics, investigations of the effects
of practice, etc. which focused on behaviors. The claim was that bistable switching
might be a kind of special behavior which affected not the outside world but perception
in the brain [57].

   Alternation of bistable images universally shares three properties, regardless of
the individual characteristics of the stimulus in use: exclusivity, randomness, and
inevitability. Exclusivity means that the two possible modes of viewing the stimulus
are never concurrently present, inevitability is the ultimate impossibility of avoiding
a perceptual reversal, and randomness refers to the length of time during which the
percept is stable on one side or the other before switching [57]. Stimuli of this type are
pivotal to this investigation, although the exact stimuli used here step a bit beyond
the traditional investigation of bistability.

1.1.3   Apparent Motion

Along the way to the visual cortex, one stream diverges from the normal pathways
of visual information, and this passes directly to the extrastriate visual cortex, to a
region called the medial temporal area (MT). MT is known to be important to the
perception of movement, and, to guiding the movement of the eyes, and to the inte-
gration of small motion percepts into a coherent picture of the moving environment
[8]. Indeed, even in patients with damaged visual cortices, perception of movement
can still be identified, even though the subjects themselves report seeing nothing.
The subjects in these cases experience what is called blindsight, a condition where
they are not conscious of any visual information, but use visual information to guide
movements or to answer questions. If pressed to say why they moved or responded
in a specific way, they will often invent answers or claim that they moved by chance
[103, 5].[5]

   When a moving stimulus is present before a viewer, the eyes collect a continuous
stream of information. When a feature of the environment moves relative to the whole,
or when the entire environment moves relative to the eyes (as when the eyes switch
foci), the retinal image is a continuous one, partially made up of all intermediate steps
between the two endpoints of motion. Micro-electrode recordings have shown that the
pathway from the eyes to V1 is well organized so that images have spatial relations
preserved in V1 neurons. Thus as objects in the environment move continuously
across our retinas, so do small areas of neuronal activation travel in a continuous
fashion across V1. Similarly, as the eyes themselves change positions, the world can

---

[5]A second interesting fact here regards what takes place when MT is damaged bilaterally. In
one case, this lesion eliminated a woman's ability to perceive motion (termed akinetopsia), and she
subsequently saw the world in a series of freeze-frames sometimes seconds apart. She would begin
pouring coffee, see the stream frozen in the air, and before her perception changed, the cup would
overflow [50].

be said to slide across the early visual areas of the brain [50].

Perception of movement however, is largely accomplished through the m-pathway, which centers in visual areas V3 and V5 (the aforementioned MT). The first of these seems to be responsible for discerning the three-dimensional shape of objects through analysis of their motion [94], while the latter discerns the direction of motion [94]. Evidence suggests that each neuron in V5 corresponds to a particular speed and direction of motion. These areas project to the parietal cortex, where information regarding moving stimuli in the environment is utilized to help guide sensory and motor systems.

It is interesting to note of this pathway that it must often make guesses about what is moving and what is not. This problem is compounded by macro- or microscopic movements of the eyes, movements of the head, and sensory blackout periods, such as during saccades (the small, fast movements of the eyes) or blinks [94]. Often only glimpses of visual information are available, but interpolating and extrapolating from them appropriately could, as this system was evolving, mean the difference between life and death.

Due to the organizational hierarchy of this system, percepts of motion can be induced without showing the eyes an actually moving object, and these are more suited to use in an EEG paradigm since in this case the simplicity of the stimuli and the specificity of their onset times are valued attributes which decrease the complexity of the collected signals.

A perception of motion can be induced in an incredibly simple paradigm, utilizing a pair of dots which disappear alternately. From this simple example, research has shown that it is the speed of oscillation and the distance between the dots (in comparison to their radii) which is able to control whether the stimuli are seen as a pair of dots flashing on and off, or as a single dot which is rapidly switching positions [1].

Most visual studies have used Stroboscopic Alternative Motion (SAM), which is similar to that used in the present experiment [96, 3, 92], but differs in that these prior investigations did not set out specifically to examine the similarities between these and prior bistable stimuli, and used a different (and less accurate) method of time-referencing.

An example most akin to the stimuli which were used in the present examination is to be found in the flashing lights around the proscenia of various theaters, or in the architecture of carnival equipment. In these cases, a band of lights is programmed in such a way that first one set of lights is illuminated, then another complementary set are illuminated and because this takes place rapidly, the illumination seems to travel from one bulb to the next. What is most interesting about these rows of lights is that as long as there are only two alternate settings, the *motion* may be perceived to be traveling in either diretion, (*e.g.* clockwise or counter-clockwise) around the pediment.

Diana Deutsch alludes to this phenomenon [25] in connection with her tritone paradox, but the visual illusion, specifically using sequential static displays of lights, is much older, dating to the beginning of the century and the research of gestalt psychology and is called *beta movement* [48]. Even at the time, it was recognized that this effect could be utilized on an industrial scale and is the basis for the optical

illusion of cinema.[6]

1.2   Auditory Bistability

The primary similarity between the visual and auditory systems is the nature of the simuli - in that both receive waves of a variety of amplitudes and frequencies. In the case of vision, these relate to luminosity and color respectively, while in the case of audition, volume and pitch.

The pressure waves in air are superimposed on top of one another, sometimes hundreds at once, and it is the task of the brain not only to deconstruct the complicated waveform into frequencies and amplitudes, but also to process these into an auditory scene.[7] Scene analysis is the process by which many sources of sound are identified and mapped in terms of their physical locations with respect to the listener, the type of sound being produced, and what these divisions likely mean. [12].

The same caution regarding the distinction between perception and sensation is certainly relevant in the auditory system as it is in the visual[8], although while the history of psychological research is riddled with multistable images and other visual illusions, the use of auditory illusions is only just beginning, facilitated by the ability of computer technology to easily and accurately create complex sounds which are tailor-made to be ambiguous to the listener. One of these in particular is called the tritone paradox, a pair of tones which can be heard either as ascending or descending depending on the key in which they are played and sometimes even just on the intention of the listener. This stimulus presents new possibilities for probing the auditory system because while the stimulus does not change, the representation does.

The history of multistability, as far as the EEG technique is concerned, has been entirely visual. While complicated auditory stimuli have been used to test auditory

---

[6]The most common example of apparent motion is before our eyes every day in the form of television. While the human eye samples the outside world at a rate of 10 - 12 samples each second [63], movie-makers present the film at a rate almost double this such that the brain cannot tell the difference between this apparent motion and real motion. Apparent motion differs from real motion in that it is a series of static images and no object is ever actually moving.

[7]*Sound* is the name given to perceptions of the pressure waves in the air. Like the light waves which serve as inputs to the visual system, pressure waves have characteristics in terms of frequency, amplitude and duration. The tiny differences between the percepts in two ears can be used to gain further information about where in the physical environment a sound originates, as can the repetitive or non-repetitive nature of sounds, and the time-coordination of pitches with different frequencies. While physicists can use powerful recording equipment and Fourier analysis to determine the frequency spectra of incoming sounds, the human inner ear does a similar job of recording using vibrating hairs, which translate the mechanical energy of the pressure wave into electrical signals to the brain, and the brain in turn performs an incredible feat of interpretation which allows some sense to be made of a complex environment [37].

[8]There are many ways in which an auditory perception, like a visual one, does not perfectly represent the environment. The precedence effect is one example of this occurrence: sound waves which reflect off of one's environment and arrive 60 - 70 msec after the initial wavefront of a stimulus will be interpreted as arriving from the same direction as the original stimulus. This allows people to hear indoor voices, for instance, as coming from finite sources, rather than as omnidirectional stimuli due to the sound ricocheting from nearby walls [73].

stream segregation in the past [4, 17, 39], to my knowledge no EEG experiments have been performed using truly bistable auditory stimuli.

1.2.1   The Auditory System

Every sound which is not a computer generated pure tone consists of many frequencies of sound, which is to say that the pressure waves reaching the ear are of many different wavelengths. A real instrument, such as a guitar (as opposed to a computer) when plucked will generate a *fundamental* frequency[9] corresponding to the length of the string, and many other frequencies at progressively lower amplitudes, but higher frequencies corresponding only in integer ratios to the fundamental frequency. For instance, while the fundamental frequency of a plucked string might be the A at 440 Hz.[10] (A440), the next loudest frequency will generally be an octave above with frequency 880 Hz. (A880 – a 2:1 ratio to the fundamental), next comes the perfect fifth, at frequency 1320 Hz (E1320 – a 3:1 relation), then another octave relation 1760 Hz (A1760 – 4:1). and 2200 Hz., etc. This sequence of progressively quieter and higher frequencies is known as the instrument's overtone series.[11,12]

When a real instrument is played, all of these frequencies sound, and their onset and offset times are perfectly synchronized which is not true of tones from separate instruments since even professional instrumentalists will deviate by milliseconds from one another. Any sound which has perfect temporal synchrony of all of its frequency components in this way is likely to be perceived as having a single source, just as, analogously, a set of line segments along the same trajectory will likely be perceived as belonging to the same line, with obstructions preventing vision of the entire thing [11].

---

[9]The fundamental frequency is recognized in this hierarchy because it is both the loudest pitch in the overtone series, and because it is the lowest (an instrument cannot physically produce vibrations at lower frequencies than its fundamental). Interestingly the information of the overtone series carries with it the location of its fundamental, so much so that when researchers played tones to subjects which were complete overtone series but lacked fundamental frequencies, subjects still reported hearing the fundamental [88, 78].

[10]One important terminological note is that musical scales (measured in units of tones), and frequency scales (measured in units of Hertz, Hz.) are not equivalent, but are instead related by a logarithmic relationship. Thus it is that as one increases by 12 semitones - the equivalent of one octave - one has doubled the frequency of the sound one is hearing. While all octave increases are the same number of semitones, they will necessarily be different numbers of Hz apart.

[11]The human auditory system can commonly hear frequencies in the range from 20 Hz. to 20,000 Hz, a staggering three orders of magnitude, although this is small in comparison to some animals such as bats. Within this range, untrained humans can consistently differentiate pitches which are separated by a fraction of a semitone, although with musical training, this ability can improve [58].

[12]Timbre is the name given to differences in the overtone series between instruments, and it is the relative amplitudes of the various components in this series which are responsible for the differences between, say, a violin and a flute, both sounding A440.

1.2.2   Auditory Stream Processing

An auditory analog to the visual illusions in Sec 1.1.2 is to be found in the auditory scene processing literature. Scene processing is the process by which sets of sounds are attributed to one or more sources in the nearby environment, just as sets of visual features are unified into coherent visual objects. Auditory stream segregation is the most well developed investigation into bistable percepts in the auditory system, and is therefore relevant background information for the present enterprise.

In his 1990 tome on auditory scene analysis, Bregman suggests that the function of perception is to take sensory input and to derive a useful representation of reality from it. The problem in the case of audition is that there is much less definite information to work with than in the case of vision, so it is remarkable that human beings can be so good at distinguishing distinct sound sources in our environment.[13]

One important concept in this field of study is that of exclusive allocation of evidence. "The exclusive allocation principle says that a sensory element should not be used in more than one description at a time," which is an important idea because this principle forms the basis of many visual illusions such as, famously, Fig. 1.3 (b), [11].

One of the most essential cues in auditory scene analysis, as mentioned above, is the timing information of when certain frequencies start and stop sounding. Humans are able to distinguish sounds presented for times on the order of μ-seconds [99], and can use incredibly accurate time-cues such as the time-difference between when specific sounds enter the two ears to help locate the source in space. When two voices or similar instruments play in unison, they will be able to coordinate extremely well with practice, but will always lack perfect cohesion. The auditory system uses these minuscule discrepancies to determine which sounds come from which instrument or voice [58].

From a neuroscientific viewpoint, this auditory scene segregation is of significant interest, since it is pivotal to the human ability to make sense of auditory percepts. Within the auditory scene, streams are objects which hold together through time, like a melody played by a certain instrument, or a sentence spoken by a single individual. Stream segregation is just as necessary for the proto-human startling at the characteristic cry of a predator, as for the modern human parsing the voices in a crowded room into separate conversations [11].

Many recent studies in this literature have used variations of the same sequence,

---

[13]Many studies have been conducted in the past half-century using separate streams of information (usually speech) presented to each ear. These studies are manifold, but largely show that attention to the sensory data from one ear allows the listener to reproduce that data, having suppressed in a way the input from the other ear. While this is certainly a notable feat, the interesting corollary is that the information from the unattended ear is largely lost, and furthermore it is not really possible for the listener to attend to both ears simultaneously although rapid switching can occur [16].

Incredibly, coherence between stimuli (in terms of their semantic content no less) has been shown [95] to overcome the directed nature of attention such that if a coherent message switches ears, subjects will likewise switch which ear's message they are repeating to follow the complete grammatical sentence. These tasks are examples of selective attention, a phenomenon which has been exhaustively studied, but still not fully understood [33].

originating in 1975 [98] of pitches which can be perceived as two types of continuous pattern. The pitch stream follows the pattern low-high-low-low-high-low, repeating, as can be seen in Fig 1.4 (a), and is grouped into either a single stream with discreet units as in Fig 1.4 (b), or as two separate streams at different pitches as in Fig 1.4 (c). Both speed of presentation and distance between the two frequencies can affect which of these percepts the listener hears [79, 39].
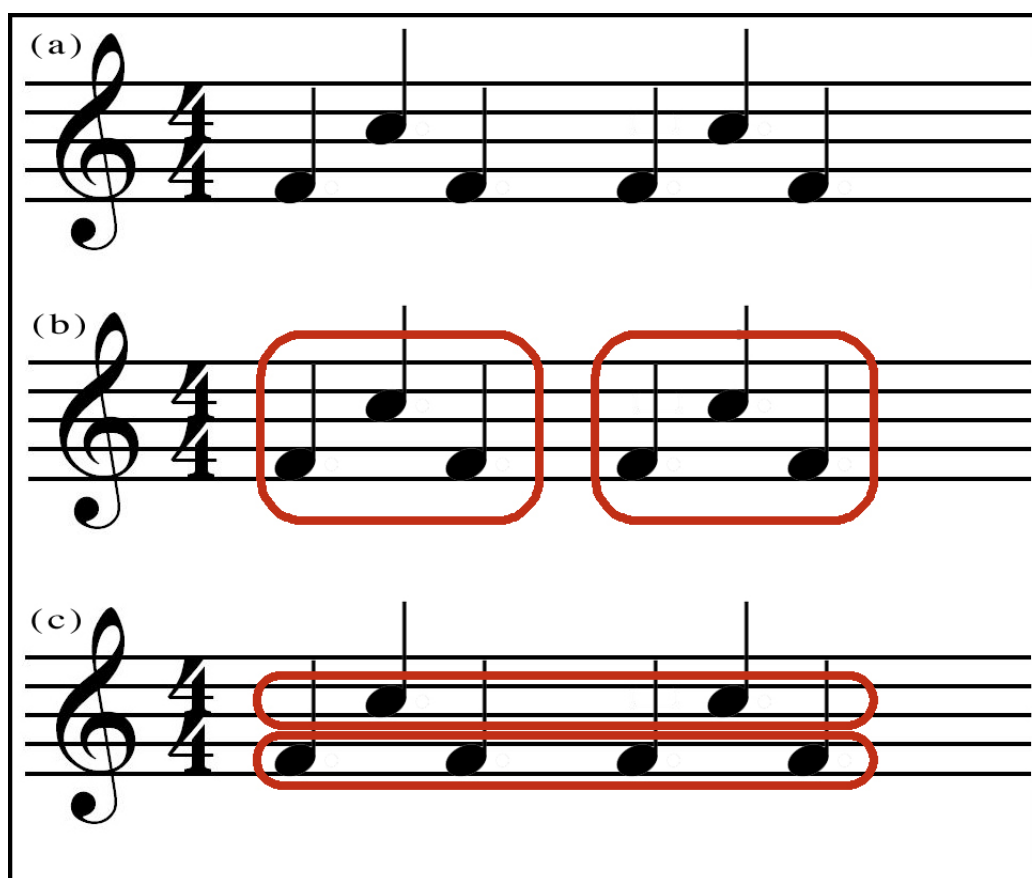
Figure 1.4: A classic auditory stream segregation task involves listening to an incoming stream of pitches at two different frequencies (a). Depending on the frequency difference between the notes, and on the speed of presentation, this stream is either heard as discreet units in a single stream (b) or as two separate streams of notes at constant pitch (c).

A 2005 study in this field by Pressnitzer and Hupé [79] highlights especially relevant and promising results. Using the classic stimulus from Fig. 1.4 to invoke bistability, the same results were sought as have previously been obtained with constant presentation of visual bistable stimuli. The results were startlingly consistent, with not only the distribution of perceptual durations (scaled by the average for each subject) remaining constant from the visual to the auditory modality, but in neither modality was the duration of a given stable period affected by the duration of the previous stable period (an indication that sensory fatigue is not responsible for the

perceptual switching). Furthermore, even when instructed to maintain a given percept as long as possible, percepts switched to a significant and comparable degree in each modality as well.

This startling similarity indicates a tentative hypothesis for the present investigation, that if an effect is to be found with auditory stimuli, it is likely to be similar to those previously isolated for bistable vision. Pressnitzer and Hupé used moving plaids as a visual comparison for their chosen auditory stimulus, and this makes sense since moving plaids can be perceived as one or two moving objects, and therefore closely match the stabilities of the auditory stimulus described above.

Interestingly however, in spite of all the diverse similarities, a difference in the magnitude of volitional control was reported between the auditory and visual domains with auditory being much more difficult than visual. While this may imply a difference in the difficulty of manipulating the two stimuli, it may also indicate that although similar architecture is present that the mechanisms responsible for bistable perception and switching in particular are to be found within the sensory systems, rather than in higher brain areas [79].

More examples of auditory bistable percepts exist, [102, 86], but as with those mentioned above, one insurmountable barrier appears when attempting to use these stimuli in an EEG paradigm.[14]. The problem is that while the brain is certainly responding to these stimuli, the size of the responses as measured in the fluctuating electric field at the scalp are very small and thus must be measured many times to attain adequate signal-to-noise ratios. For this to be achieved, the stimuli must be presented many times, and a continuous stimulus, like that of the auditory scene analysis studies will not serve. We look therefore to the literature on musical illusions for a stimulus which is not only bistable, but which has a well-defined onset, and which takes a short time to present.[15]

---

[14]The McGurk effect is an example of a multimodal illusion (although not a bistable one) in which visual information and auditory information about a spoken syllable combine such that a unique syllable is perceived which does not match either of the original stimuli [61] In an innovative study, this illusion was reworked into a bistable paradigm by using a computer generated face instead of a real one to speak the syllables. The face was seen in profile, and was seen speaking to another person with a similar profile across from it.

The outline of the two faces delineated the edges of what seemed to be a granite vase. As the computer generated video progressed, either one of the two faces was perceived to move its lips to form a simple syllable /aba/, or the vase was seen to rotate such that a chip in its stone surface was shown in profile, combining the McGurk paradigm with that of the Rubin vase. This brilliant manipulation allowed the viewer to either interpret the visual information as a pair of faces (which hold speech information) or as a stone vase, (which does not). The face-vase percept is known to be bistable, and the researchers found that when participants perceived a vase the McGurk effect was absent and while when a face was perceived, the McGurk effect was in play indicating that the illusion is not a sensory one, but a consequence of some higher integration of semantic information [66].

[15]In the case of auditory illusions, the various cues upon which we form our judgments of a sound are somewhat different from but no less varied than their visual counterparts. The basic elements of human perception of music are: pitch, timbre, volume, tempo, and rhythm. These can be set at odds in a number of ways. Say, for instance that the sound of an engine was to pass us by, undergoing a standard doppler shift downward in pitch as it did so, but say furthermore that the sound of the engine was to grow softer as it approached, and then louder again as it sped away, which perceptual

1.2.3 Shepard Tones

A rich category of auditory (and arguably *musical*) illusions are based on a constructed tone called a Shepard tone, first constructed in 1964 by Roger Shepard, a researcher at Bell Laboratories. Shepard was experimenting with the difference between two important concepts in music theory: *pitch height* and *pitch class* [89].

Musical pitch height is a linear dimension, corresponding by a logarithmic relationship to the frequency of the sound being played. When one sound is said to be higher or lower than another, pitch height is being described.[16] To hear the linear nature of pitch, one need only sweep one's fingers over a piano keyboard and note that the sound consistently rises or falls and does not repeat or alter its rate of change [28].

Pitch class, by contrast, is a circular dimension which repeats every twelve half-steps in Western music. In Western tonal music, unison and octave intervals are thought to be harmonically interchangeable and chord inversions (any rearrangement of the pitches vertically) are considered equivalent to their parent chords [74]. Twelve half-steps comprise an octave (a precise doubling of the frequency) and indeed, it is not only the Western chromatic scale which seems to be bound by octave intervals. Nearly every culture in the world incorporates some form of circular dimension repeating at octave intervals [19].

The question then arises: at what pitch height, and into what pitch class do we put a tone which is comprised of many frequencies? The question is answered by the fact that all sounds from our environments: thunder, violins, speech, etc. are made up of many frequencies, but always characterized by a single "fundamental" frequency.[17] This is both the lowest, and characteristically the loudest frequency of the overtone series, and is therefore the one which identifies the pitch of the sound.

The dimensions of pitch class and pitch height were for a long time thought to be orthogonal. One can visualize this by imagining the two dimensions as angle and height coordinates on a helical arrangements of notes as in Fig. 1.6. If one looks at only the height of the notes, they increase linearly, but if one examines the diagram by looking along the vertical axis from above, then the notes appear to circle around and around. In a paraphrase of Deutsch's words, the orthogonality of pitch class and

---

cue are we to trust, did the engine approach us as the pitch would indicate or did it move away and come back as the volume signaled? Another example would be, for instance, a voice speaking agitated words in a slow calm voice. In this case, the semantic content of the sentences is at odds with rhythm, tempo and timbre, all of which are used in speech to indicate an emotional color for our sentences [67].

[16]Interestingly, the conventional *direction* of pitch height is merely that, a convention, and many musicians would be surprised to learn that the ancient Greeks thought in the opposite way, with notes of longer *wavelength*, (lower pitch and frequency) being thought of as sounding "higher" [58].

[17]Any real instrument plays not a single note, but as noted in Sec. 1.2.1 an immense, diminishing series of pitches above a *fundamental* frequency, all of them sounding at once to make up the unique sound of that instrument. When we speak of A440, played by the first violin of an orchestra so the other chairs may tune their instruments, we refer not to the complete set of pitches which are actually sounding, but to this fundamental frequency and not to the overtones. This is the answer for the majority of instruments, and no confusion arises as long as the fundamental is the lowest and loudest of present sounds.

perceived height is a concept which is generally taken as axiomatic. If a musician were faced with the question "Which note is higher, C-sharp or G?," he would probably reply that the question were nonsensical: one would have to know; *which* C-sharp and *which* G before a meaningful answer could be given [20, 19].

Shepard's investigation then, was into the nature of this orthogonality and like any good scientist, he set out to find an exception. Shepard designed a tone which would convey information of pitch class only, not pitch height. His tones consisted of 10 octave-related sinusoidal sound waves (it has been shown that sawtooth and square wave shapes induce the same effect [36]), whose amplitudes were constrained by a fixed, gaussian envelope. He predicted that due to the fixed nature of the envelope, the perceived heights of the tones would remain invariant despite differences in pitch class [89].

Shepard's primary finding for the purposes of the present investigation was that when two such tones of different pitch class were played one after another, people judged the direction of motion (on the scale of pitch height) based on proximity. That is, the jump from one pitch class to another can be achieved in either direction around the pitch class circle, and listeners were able to hear the relative arc-lengths of these two paths. For example, if the first Shepard tone were built around pitch class C, the arc-length to pitch class D would be shorter than that to pitch class F (see Fig. 1.6).[18]

One possible use of this finding is the creation of an auditory analog of Escher's print "Ascending and Descending" which is an artistic rendering of a staircase which is seen endlessly looping back on itself. This is an example of the Penrose Steps, which function by breaking normal lines of perspective [72]. This is accomplished by playing successive Shepard tones around the circle. The overall sound never gets lower, (in fact it will repeat every 12 tones), but will feel as though it is always descending.

1.2.4   The Tritone Paradox

Shepard's research was of interest to engineers, musicologists and psychologists alike, but as so often happens, he ignored the most interesting loose ends in his data. While people hearing pairs of his tones made judgments of pitch height based on proximity, he did not investigate the tantalizing question of what happened when he also removed the proximity information as well.

The tritone paradox is the name given to a musical paradox discovered in 1986 by Diana Deutsch. The paradox is made up of a pair of successively presented tones, each of which is composed in the manner of Shepard's tones above. Each frequency in the first tone falls into a single pitch class, as does each frequency in the second

---

[18]One consequence here is that one's ears can be led around and around the pitch class circle indefinitely without the actual sounds changing with the disconcerting result that the sounds in one's ears appear to be rising or falling infinitely without changing octave. A better known visual analog of this illusion is the M. C. Escher lithograph print: "Ascending and Descending [15]," seen in Fig. 1.5.
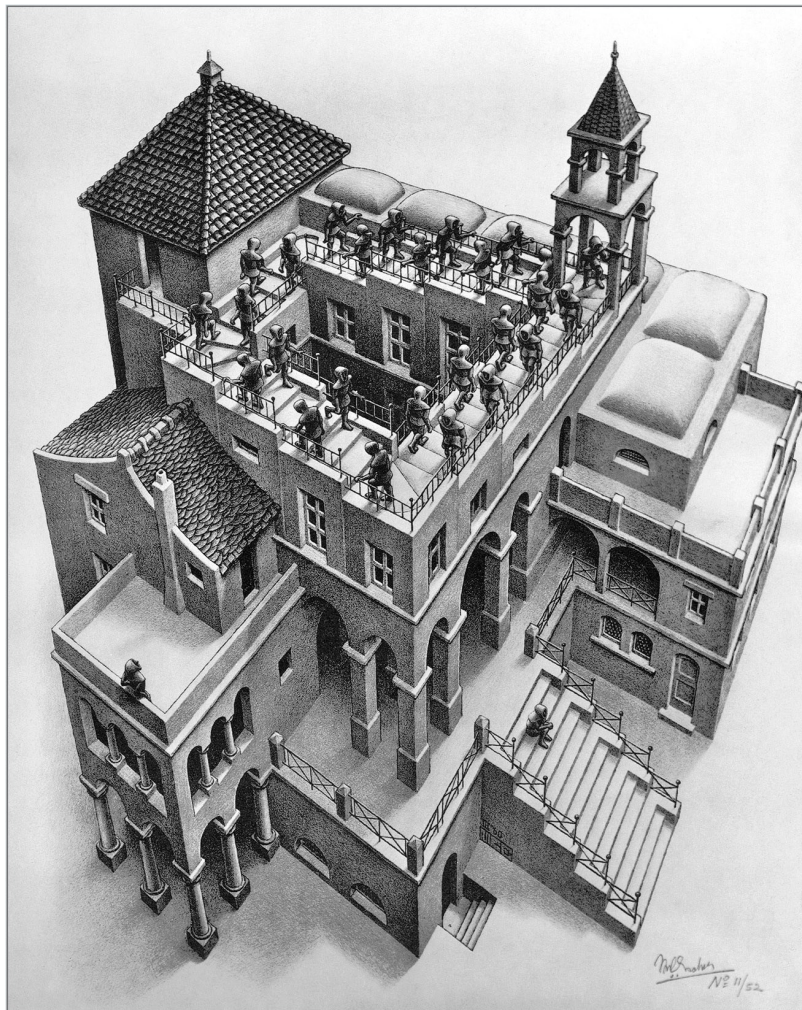
Figure 1.5: This Escher print makes use of the Penrose Steps [72], an illusion which functions by breaking traditional rules of perspective drawing, to create an endlessly descending staircase.

tone. The two pitch classes employed for this specific musical illusion are those on opposite sides of the pitch class circle and are therefore related by an interval of a diminished 5th which is equivalent to six half-steps and is called a tritone.[19]

---

[19]In the design of the Western, well-tempered piano, the notes are struck by hammers, positioned at an exact and consistent distance down the string being struck. When a musical instrument is plucked, struck, bowed, etc. in this manner, a node is created at the point of contact, and certain frequencies are inevitably damped and eliminated. The piano is designed so that every string is damped to silence the tritone from the overtone sequence. In most Western music, this tone is thought to be extremely dissonant (even evil) and therefore was systematically removed.
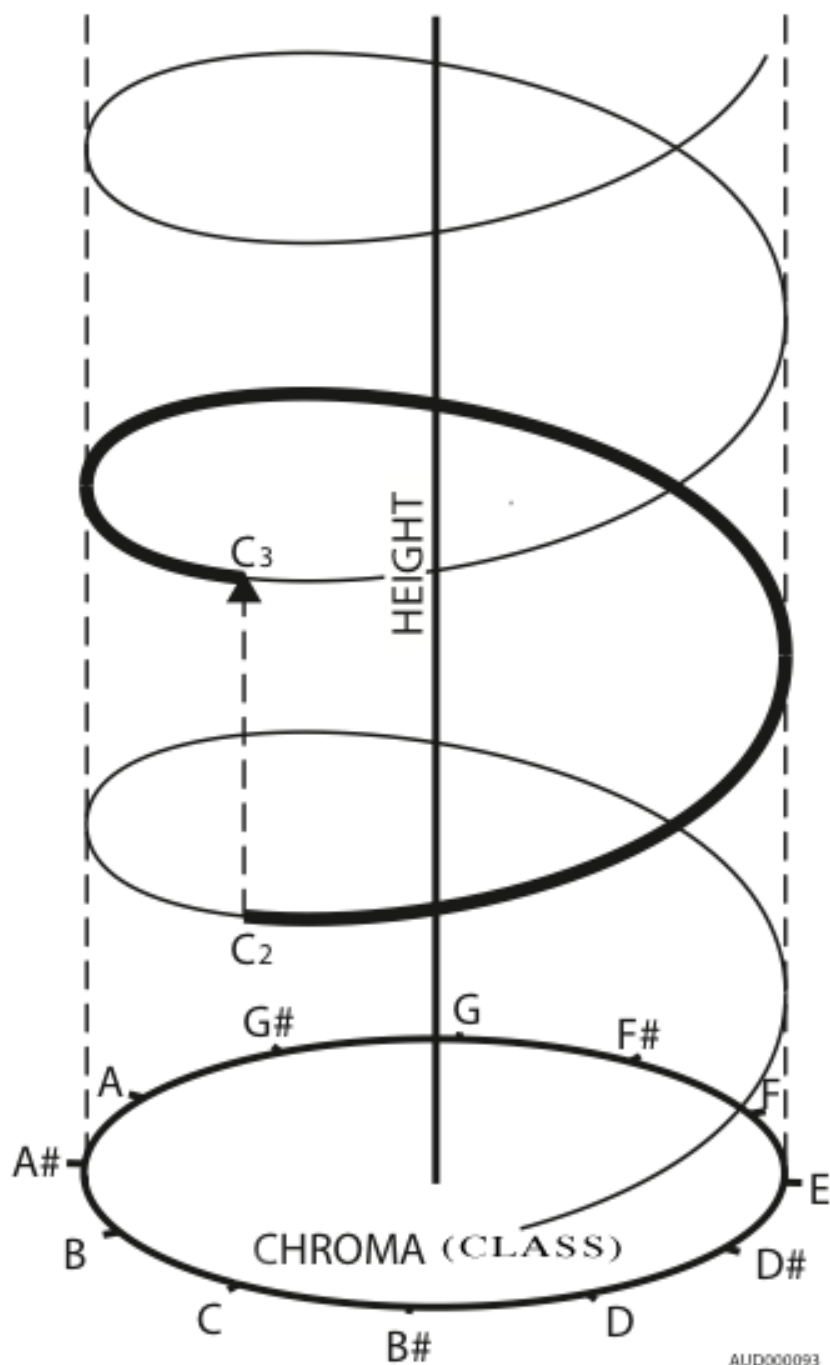
Figure 1.6: The orthogonality of pitch height (shown here as position along the helix's axis), and pitch class, (shown here as angular position). All elements of a given pitch class (exemplified here with pitch class C) are at the same angular position. Image borrowed from Shepard 1982 [90].

When individuals, regardless of musical training, listen to these successive tones, and are asked to judge which is higher than the other, they cannot use the standard cue of fundamental frequency of the sounds, nor can they use the secondary cue of proximity on the pitch class circle. The result is a fascinating paradox: some will hear the second tone higher than the first, and some will hear the first tone higher than the second. Furthermore, as the key of the pair of pitches ascends up the scale, the direction of movement from the first pitch to the second can alter for the same individual. Thus, as Diana Deutsch wrote on first publishing these results:

"When played in one key it is heard as ascending, yet when played in a different key it is heard as descending instead. When a tape recording is made of this pattern, and it is played back at different speeds, the pattern is heard either as ascending or as descending depending on the speed of playback. To add to the paradox, the pattern in any given key is heard as ascending by some listeners, but as descending by others" [18]. She continues, with respect to the implications of her work: "The notion that a melodic pattern might be perceived as radically different under transposition appears as paradoxical as the notion that a visual shape might undergo a metamorphosis through being shifted to a different location in space [18]."[20]

Starting in 1986, Diana Deutsch researched causes and effects related to the tritone paradox, shedding light on the systems of human cognition, audition and on hereditary traits. Her initial study looked only at musically trained volunteers, used a small sample size, and used stimuli which were limited to a central range of frequencies, matching human vocal production [18, 21, 28].

Within a year of her first work on the subject Deutsch had corrected the several immediate limitations of her original investigation. In a technically advanced paper, published now in a journal of psychophysics, she sought to understand how the specific logarithmic frequency band containing her stimuli might affect the outcome of her original experiment. After using a larger number of participants, and using four different spectral envelopes for her stimuli, set at half-octave intervals, Deutsch concluded that although some results were changed in small ways, such as the total number of pitches heard descending, the overall pattern of results was preserved regardless of what frequency band was used [19].

In the same year, another study was also conducted to test the responses of persons without musical training. Participants with and without musical training (defined as two or more years of training) were compared, and no differences were found [30]. Deutsch then set out, over the course of the next decade, to discover what the causes were of the widespread and striking individual differences in listeners interpretation of the tritone paradox.[21]

---

[20]And the ramifications could be quite staggering musically. Not only is the complete orthogonality of the two dimensions of pitch disproven, but it would seem that regardless of their level of musical training, most individuals have some ability to hear absolute pitch.

[21]Given that one's musical experience did not affect the way individuals heard the tritone paradox, the logical next question was to ask what did affect it.

The first breakthrough came in 1990, when Deutsch found in speech what she had sought in music. The data showed that rather than correlating with musical training, each individual's responses to the tritone paradox corresponded instead to his or her own unique vocal range [31].

These early results of Deutesch's work are relevant to the present endeavor, in that they demonstrate how robust the unexpected responses to the tritone paradox can be, such that neither musical training nor logarithmic frequency band has any large effect.

Deutsch, like Shepard before her, explored a liminal region of a previously known system, and discovered a promising area for research. A further liminality existed in her data however, the ambiguous keys (one or two of which existed for each subject) which indicated an equal number of upward and downward perceptions of the tritone paradox stimuli. That is, for each subject, there are one or two pairs of tones that are equally ambiguous – that are heard as ascending or descending in equal measure. These ambiguous keys are exploited in the present experiment as bistable auditory stimuli, used in conjunction with EEG data collection to explore the boundary between auditory sensation and perception.

Indeed in a later article, Deutsch publishes strings of musical notes which would be heard in completely different ways by different listeners.[22,23,24]

While correspondence with vocal range is not a difficult correlation to accept, Deutsch also found that persons living in California and in England perceive the same stimuli in systematically opposite ways. Some individuals do not match this trend (notably her original eight subjects fall all along the scale, and are all Californian), but after enough subjects have been averaged, a given region seems to exhibit a specific typical response to the tritone paradox [21].

A final piece of this puzzle, as far as Deutsch's research directions are concerned, focuses on the heritability of typical responses to the paradox. In several populations, not only do parents and their children have similar perceptions of the paradox, but this dimension interacts with their place of origin, such that even within small communities children of parents who were raised locally and who were not, form statistically distinct groupings of responses [82, 23, 82, 24].

[22]Perhaps the most sobering result of Deutsch's research is that melodies may not be singular objects, that is, they may sound differently to different people every day. As Deutsch herself put it: "the intriguing possibility has now arisen, for computer- synthesized tones at least, of producing music that not only sounds quite different under transposition, but also sounds radically different from one member of an audience to another, [19]." Although it is unlikely that this type of illusion actually does occur in normally composed music, an enterprising composer could take advantage of the effect to create a whole new type of audience experience [28]

[23]Claud Risset, a french composer has used tones such as those created by Shepard to create music which disobeys traditional laws. Listeners may hear tones which rise or fall endlessly without seeming ever to get any higher or lower. This is due to a clever exploitation of the same principle, in which tones loop the pitch class circle at many different octaves simultaneously without allowing the power in any given frequency band to alter significantly [25, 84].

[24]One fantastical consequence of the results of tritone paradox studies is that nearly every person, regardless of background or musical training, has a form of perfect pitch. This is evident from the fact that regardless of the absence of pitch height information, these individuals are able to tell what pitch class they are hearing (as evidenced by consistent responses). Theories now vary as to why this is the case, but it is an interesting phenomenon nonetheless. This form of absolute pitch differs significantly from what is commonly called "perfect pitch," which is an explicit ability to identify heard pitches [101].

1.3   EEG studies of Bistability

One established procedure for investigating sensory systems is electroencephalography (EEG), which is the non-invasive recording of neuroelectrical activity through electrodes contacting the scalp. Participants in EEG studies wear cloth caps with electrodes sewn in to create a rough spatial map of voltage changes in populations of neurons over time.

Figure 1.7: The Experimenter wearing the EEG cap at Reed College's Sensation, Cognition, Attention, Language & Consciousness EEG Laboratory, acting as a subject in his own experiment.

Participants are presented with repetitive stimuli (via any sensory system), and their neural responses over a large number of trials are averaged to reduce the contribution of brain activity unrelated to the stimulus or task (often referred to as noise) by increasing the signal to noise ratio once these trials are averaged. While the individual signals from the brain are unnoticeable to the naked eye in the raw signal, the EEG, averaged across many trials, allows the noise to cancel out and shows the *event related potential* (ERP), which is the neural response specifically related to a given stimulus or response. At the location of each electrode accurate temporal information can be recorded as to the characteristic response of underlying brain structures to the specific stimulus but it is impossible to prove any particular correspondence between a scalp distribution of electricity and the specific sources inside the brain because

multiple arrangements of sources can give rise to each possible surface distribution. This is known in physics as the *inverse problem*. The best that can be done is to mathematically calculate likely distributions and to make educated guesses. The EEG technique can be taken further by constructing more elaborate stimuli which involve multiple sensory modalities, which are presented only to one ear, eye or hand, which require active response from the participant, or which are accompanied by instruction about direction of attention and so on [60].

The EEG technique is useful because the strength and timing of the electrical responses recorded at various scalp regions can reveal important aspects about how incoming stimuli are encoded, processed, and understood under different experimental conditions. The incredible temporal resolution of this procedure distinguishes it among other neuroimaging methods. Functional Magnetic Resonance Imaging (fMRI) can show spatial resolution on the order of millimeters, while EEG responses, even with 96 or more electrode sites, are only accurate to within a couple of centimeters, (a figure which is itself debated as some would claim EEG recordings have no spatial resolution at all), but this 20-fold loss in terms of spatial resolution is accompanied by a several-thousand-fold improvement in temporal resolution over fMRI. EEG can record events on the order of milliseconds or faster, while fMRI can image the brain only every few seconds. Fig. 1.8 shows the relationship between the resolutions of these and other neuroimaging techniques in spatial versus temporal dimensions.

In most experiments, to answer a question using an event-related potential (ERP), a researcher will show two images to a subject which differ in some important way, *i.e.*, one of which is a face, and one of which is a house. The researcher repeatedly shows these images to the participant, and upon analyzing the data finds that the brain responds differently to a face than it does to a house. The problem with this approach is that it does not control for physical features of the image which is being shown. Since the two images were different in the first place, of course they can be expected to be processed differently in the brain and there is no way to tell whether these differences reflect sensory, perceptual, or cognitive levels of processing. As it turns out, faces and houses are processed differently, but countless EEG experiments ignore this type of confound, relegating it to the realm of issues which can not be avoided [10, 9]. Bistable stimuli address this problem handily.

1.3.1   Necessary Characteristics of EEG Stimuli

Experiments employing EEG incur certain limitations in terms of their characteristic stimuli. One of these is that the crucial aspect of the stimuli must be replicable some hundreds of times. Usually this means the stimuli will all be identical, but some language studies use sets of words which share a characteristic (four-letter nouns) but all of which are different. The ERPs are often such small signals that more than 100 trials are necessary to find a consistent effect among the noise in the signal. This characteristic makes some studies difficult to undertake since a stimulus which takes 5s to present will easily lead to participation times of many hours, and in addition to higher compensation costs, subject fatigue and learning become important factors.

Stimuli must have a well-defined onset time against which to time-lock the EEG segments to be analyzed. While continuous stimuli, or stimuli which are distributed in time are acceptable and even desirable for hemodynamic studies using fMRI, electrophysiological techniques such as EEG need as much temporal specificity as possible in the stimulus so as to preserve the same in the neural response.

EEG stimuli must also be fairly simple for the straightforward reason that ERPs are not clearcut results. For no stimulus will a perfectly clear electrical component be discernible, but the simpler the stimulation, the more straightforward is the task of parsing the results. This is the reason that colored shapes, simple geometric solids, pure tones, clicks, etc. are preferred to more complicated stimuli.

These limitations are challenging, but are not ultimately prohibitive. Some creativity and research is often required to discover exactly the correct stimulus.

1.3.2   Exploring Novel Stimuli for EEG

Until the present investigation, the stimuli of the tritone paradox have not been discussed as bistable sounds, but just as Diana Deutsch examined Shepard's data and found an ambiguous center between unambiguous extremes, there is an ambiguous center to be found within the tritone paradox as well. While each subject consistently displays regions in which it is highly probable or highly improbable that the tones will be heard descending, there is nearly always a point between these regions where a descending percept is equally as likely as an ascending one. Deutsch makes only one reference to this ambiguous, liminal state in her published papers on the topic, in a figure caption in her research summary in Scientific American [28]. It is important to note that when the pitches were heard descending in half of the trials, this is not to be interpreted to mean that subjects could not tell and were guessing, but rather that in half of the trials, (depending perhaps on what tones were presented previously, *i.e.* a priming effect), a completely stable descent was perceived and vice versa.

In a 1999 review of prior literature and a proposal of a new mechanism for bistable alternation, Leopold and Logothetis codify three ubiquitous features of multistable perception: exclusivity, inevitability and randomness [57]. The first of these is, in theory, a consequence of the sensory system, and is supported in the case of the tritone paradox by the unambiguity reported by Deutsch's subjects. When asked to make a judgment about whether the sounds rose or fell, participants were easily able to decide, and rarely reported some combination or exception. Inevitability is trickier, and since the tritone paradox has not previously been utilized in this way, no literature exists on the subject. From pretests and subject reports, the percept was said to reverse at times without intention, but not to the degree common in static, visual bistable figures, such as the Necker cube. Voluntary control of switching is possible for human listeners, but again no numerical data exist to further these qualitative reports. In one previous comparison of auditory and visual bistability [80], auditory switches happened inevitably and with a similar temporal distribution (though less commonly) as visual. Finally, an analysis of randomness in reversals of the tritone paradox also suffers from the lack of a literature on the subject, but again participant reports indicate no discernible pattern to the periods of stability and reversal.

A similarly random and inevitable distribution of behavioral data would indicate a striking similarity in the processing of bistable auditory stimuli, and would support a top-down hypothesis (*i.e.* some involvement of intentional control) of perceptual switching which occurs at a level beyond that of cross-modal integration, whereas a different (non-random, non-inevitable etc.) distribution would support sensory models of bistable perception.[25]

A visual stimulus was designed for the present investigation which is meant to be an analog of the tritone paradox in as many ways as possible. Like the auditory stimulus, this visual illusion comes in two pieces (frames), and relies on scene analysis to convey a perception of a rotating wheel when no piece of the image was actually moving. Here the data from Pressnitzer 2006 [80] are more relevant, although still, that prior investigation used continuous stimuli rather than discreet, repetitive ones. Similar stimuli have been used previously, although these were simpler than the image in the present investigation. These prior stimuli are known as Stroboscopic Alternative Motion (SAM), [96, 3, 92]. As noted previously, however, these prior investigations referenced their analyses to the participant's manual responses, rather than to the stimulus onset and therefore acquired less certain time-stamps for the components they reported.

1.3.3   ERP Components Common to Bistable Stimuli

Many ERP components are associated with the onset of a visual or auditory stimulus, most of which occur soon after the stimulus is presented. Several characteristics are often reported in studies of bistable stimuli, and these are a good starting point for the present investigation.

To date no EEG studies have been performed using auditory bistable stimuli, although a few single-unit and fMRI studies have been conducted. For example single-unit recordings from the auditory cortices of monkeys showed a similar probability distribution in response to a set of ambiguous sounds as that of ten humans tested previously with non-invasive procedures, meaning that results were correlated at the behavioral and neuronal levels for this brain structure, which is therefore likely to be the neural locus of the effect [64]. Separate fMRI studies of perceptual reversals implicated the intraparietal sulcus [17] and the medial temporal gyrus [47], among other regions of the brain.[26]

---

[25]Reversal rates are likely controlled by a complex of neural networks which work together to allow increased rate, decreased rate or more complicated patterning. "Recently, transcranial magnetic stimulation (TMS) has been employed to test the causal role of frontal-parietal areas in initiating perceptual reversals [44, 104]. Interestingly, disruption of activity via TMS in different subregions of the parietal cortex appears to result in opposite effects (increasing or decreasing reversal rates), thus suggesting that a more complex network of parietal regions are involved in bistable perception [45]." [14].

[26]Past studies looking at bistable auditory stimuli (drawn both from the literature on auditory streaming and on the McGurk effect) have examined the possible loci of differences between percepts. An fMRI investigation of an ambiguous /ada/ versus /aba/ phoneme found differences between percepts in the planum temporale, and in the superior temporal gyrus and sulcus, as well as in the medial temporal gyrus [47].

It is unclear as yet what specific electrical results can be expected for auditory stimuli, and the best tack may be a comparison to the literature on visual bistable stimuli.

Reversal Negativity

Electrophysiologically, there is a characteristic potential associated with the comparison of bistable percepts which alter from trial to trial, and those which remain constant. Called a reversal negativity (RN), this component has not been fully explored, and is in part the subject of this investigation. It is not clear whether this component is a positivity associated with the brain's attempt to sustain an existing percept or a negativity, associated with the novelty of a new percept [76]. This component is reported to be "maximal over parietal-occipital scalp regions, begins at 170 ms after the stimulus, peaks at 250 ms, and persists until 350 ms. [76]." It is likely that these signals are generated in the ventral occipital-temporal cortex [76], and this result is consistent with recent fMRI studies employing similar stimuli and paradigms [42, 49].[27]

Late Positive Complex

Another previously reported signature of bistable reversals is a late positive complex [62] (LPC), which is likely based in the intraparietal sulcus (IPS) and surrounding superior parietal lobes (SPL) bilaterally [65]. It is thought that this component does not depend on reversals per se, but is more closely related to processing of updating short term memory [76].

While it is not surprising that a set of visual stimuli may produce differences in the posterior temporal and parietal areas, it is possible that differences in frontal regions may be present as well. The intention and effort associated with perceptual switching or preventing a perceptual switch may arise, especially in the case of a difficult switching task, as novel components in the ERPs.

In another study, ten humans were pretests with the auditory streaming lo-hi-lo stimuli (Fig. 1.4), and a probability distribution was constructed showing how often they heard the sounds grouped in each way dependent on physical features of the tone sequence, such as speed and frequency difference. Subsequently two awake rhesus monkeys were tested with micro-electrodes located in A1 in their auditory cortices. The results showed a pattern statistically similar to the probability distribution from the ten humans, showing again that a difference between percepts was present in this area [64]

Yet another fMRI study implicated the intraparietal sulcus, but based on the background research which supported this investigation, it is likely that this area was more related to object identification and stream segregation than the separation of bistable stimuli [17].

Finally, a study conducted with Magneto-encephalography (MEG) found cortical differences between streams, as well as differences in the auditory cortex, so a possible later component, originating cortically is possibly implicated in bistable processing [39].

[27]In many, but not all studies of visual bistable figures, a component called the Reversal Positivity (RP) occurs around 130 ms after stimulus onset, with a peak width of $\pm$ 35 ms. This component is most easily visible at occipital electrode sites, and occurs only when the reversal was induced endogenously [55]. Other studies have also reported this component, finding it to be of amplitude at or below 1 μV [53, 54]. This is the earliest known reversal-related ERP component, and is one way of differentiating responses to endogenous and exogenous reversals.

Interestingly, The LPC is found for reversals whether they were endogenously (spontaneously) created or exogenously created (by showing a stimulus unambiguously one way and then the other way) [69]. This indicates that this component is largely unrelated to the intention to switch, or even to the causation of the switching through some non-conscious mechanism and indeed it has been suggested [92] that the LPC is an index of conscious realization of the reversal.

1.3.4 Source Localization & The Inverse Problem

Historically the spatial resolution of EEG was far inferior to that of hemodynamic techniques such as fMRI. This was due to the inability to measure electric field strength at any point other than the surface of the head, and also to the inverse problem, which states that it is impossible, knowing only the field strength at the surface of a body, to prove what sources gave rise to that field. There are in fact an infinite number of possible source distributions, but through the use of simplified models, knowledge of the boundaries of regions inside the brain, or Montecarlo statistical formation, educated guesses can be made.

1.3.5 Hypothesis

It is proposed that ERPs will be collected which show common characteristics between visual and auditory perception of bistability. Furthermore on trials on which perception switched relative to the previous trial, it is hypothesized that components will be discernible which match those previously isolated for static visual bistable stimuli. In the pre-stimulus interval, a previously unexplored region, new ERP components may become apparent and provoke further investigation and theory. The null hypothesis in this case is that these ERPs will not significantly differ between reversal and stability conditions; that is, there will be no characteristic components associated with perceptual switching of dynamic stimuli.
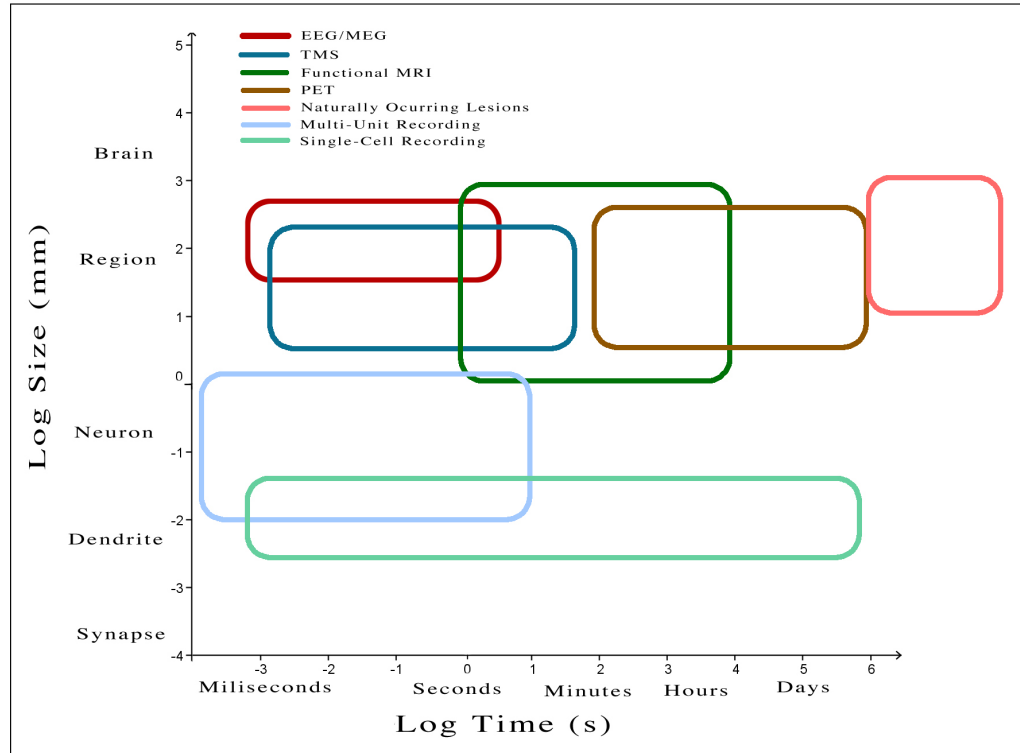
Figure 1.8: A now classic graph, showing the relative resolutions in space and time of the myriad of modern neuro-imaging techniques. Notably, fMRI can acquire more accurate spatial data than can EEG (these two are the most common), but EEG can record far more accurate timing data than fMRI. This figure is useful for understanding the relationships between techniques, but does not tell the entire story – single-cell recording and multi-unit recording, for example, are able to record far more accurate spatial and temporal data than other techniques, but these are invasive techniques and are not used on healthy human subjects. While it may seem that higher resolutions in both dimensions (the bottom left quadrant) would indicate superior techniques, those measures which are capable of tracking changes across the whole brain or across a lifetime are equally essential to a holistic understanding of neural function.

Chapter 2

Methods, Measures & Procedure

As with every research technique, EEG sports scientific strengths and weaknesses as well as procedural conveniences and challenges. While a small sample size is usually sufficient to find statistical results, lengthy preparation times make participation a significant time commitment, and while contemporary computer software makes data collection an automated process, reducing noise by improving electrode-scalp connections is arduous and somewhat unpleasant for the research subject.

In this chapter, procedures generic to all EEG experimentation, and particular to this investigation are detailed and recorded.

2.1 Participants

Participants in this study were healthy adults, all of whom were students enrolled at Reed College (36% female, mean age = 20 yrs old, SD = 1.5 yrs). Participants' electrophysiological responses change as they age [34], and so an upper bound was imposed on participant age of 30 years. There were two criteria for exclusion of otherwise qualified subjects. These populations were persons reporting a history of brain injury or any other neurological condition that may affect their electrical brain activity and persons with uncorrected sensory (visual or auditory) deficits. None of the individuals who expressed interest in participation were excluded based on these criteria.

The total number of participants was 18, enough to secure significant results if they were to be found. Data were collected from January 2012 to March 2012.

For participation in the study, participants received five entries into a lottery, hosted by the Reed College Psychology Department. A winner was drawn for this lottery at the end of the spring semester, and the winner was to receive $150.00. Additionally, several participants were younger members of Reed's SCALP-EEG lab, and participated in the experiment as part of a training program to work in the lab.

Each participant was assigned a number code, and other than submission to the psych lottery, all further data processing was done via these codes, not via participant names. Names were collected so that participants might be entered into the psych lottery, but these names were kept separately from all data and no information was

kept which could reconnect them.

2.2 Criteria for Exclusion

Of the 18 subjects originally tested, one was excluded because the presentation code contained an error at the time of testing, so the ISI were shorter in one condition than the other, rendering his data useless. Another subject was excluded for responding in an improper manner which demonstrated that the task was not properly understood. Two further subjects were excluded due to a small number of trials in one condition (Auditory Reversals), with each subject hearing only about 60 reversals in 600 trials. The minimum number of trials/condition for inclusion was set at 120, and all other subjects achieved this in each condition, even after artifact rejection. This left 14 subjects for inclusion in the final analysis.

2.3 Stimuli

Stimuli in both modalities were novel bistable stimuli which were believed to be similar to each other in as many aspects as possible. Since presentation speed and the length of the inter-stimulus interval (ISI) are known to be important factors influencing reversal rates, [70, 71], (or [56] for a figure) these were kept constant across conditions and were maintained in a range known to produce bistable perception of static, ambiguous, visual stimuli.

Some prior investigations [69] used an ISI of 3300 ms, which more recently has been suggested [55] might easily result in participants experiencing separate percepts, rather than perceptual reversals. Thus such a long ISI was avoided in the present study.

2.3.1 Auditory Stimuli

The audio stimuli for this experiment were the same as those created by Diana Deutsch, and were downloaded from her website [27]. Four tone pairs were downloaded and were transposed to create the remainder of the chromatic spectrum. At the time of this transposition, it was thought that Deutsch had used a different spectral envelope for each stimulus, but it was later discovered that this was not the case, and therefore the stimuli used in this investigation were created under several spectral envelopes.

The tones were each stationary Shepard tones, consisting of six octave-related harmonics whose amplitudes were bounded by a gaussian spectral envelope. The tones' octave relationships required that all frequencies comprising each tone be drawn from the same pitch class. The pitch classes for the first and second tones differed by exactly six semi-tones, an interval known in music theory as a tritone. As mentioned in Sec 1.2.4, the tritone is meritorious for this purpose because it is equally distant (in terms of semitones) from pitches of the same pitch class upward and downward along the scale. A fixation cross (a "+" at the center of the screen) was left in place on the

screen during the entire testing period regardless of stimulus modality. A schematic of the spectral breakdown of these tones is shown in Fig. 2.1[1]
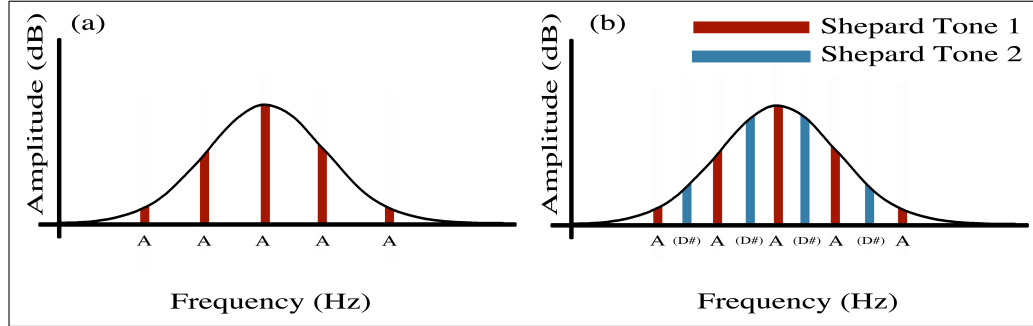
Figure 2.1: Part (a) of this image shows the arrangement of a Shephard tone, with frequencies related by octaves at varying amplitudes. Part (b) shows the relationship between the two Shepard tones used as audio stimuli in this investigation. In the image, these are represented by pitch classes A and D#, but for a given subject, these might be replaced by any tritone-related tone pair.

Although this investigation restricted its stimuli and auditory pretest to 12 pitches, this is merely a convention of Western music, and auditory stimuli could be created in such a way that a range of frequencies more sensitive than the twelve semitones of the Western chromatic scale could be used.

2.3.2 Visual Stimuli

Visual stimuli were created using Presentation's vector graphic generator, and consisted of 12 circular discs of radius 0.79 cm (30 px) arranged at equal intervals around the perimeter of a larger circle of radius 6.43 cm (243 px) at the center of the screen. The central circle was not visible, but the discs were connected by a thin line running through their centers. The entire image therefore traced a visual arc of 9.8°, given that the subject did not move his or her head appreciably during testing. A fixation cross (a "+" at the center of the screen) was left in place on the screen during the entire visual trial.

The two stimuli can be examined in Fig. 2.2 (a) and (b), and it should be noted that the positions of the discs in 2.2 (b) are halfway between those in 2.2 (a). The design of the visual stimulus was intended to mimic its auditory counterpart as closely as possible.[2]

Visual stimuli of this type have been previously shown to elicit strong percepts of motion in viewers, and the distance between the discs (a product of their individ-

---

[1]Examples of the auditory stimuli used by Deutsch and in the present investigation are available online at: $http://philomel.com/mp3/musical_illusions/Tritone_paradox.mp3$.

[2]A video, hosted by YouTube of the experimental stimuli is available at $http://www.youtube.com/watch?v = pMyVmYVUav4$.
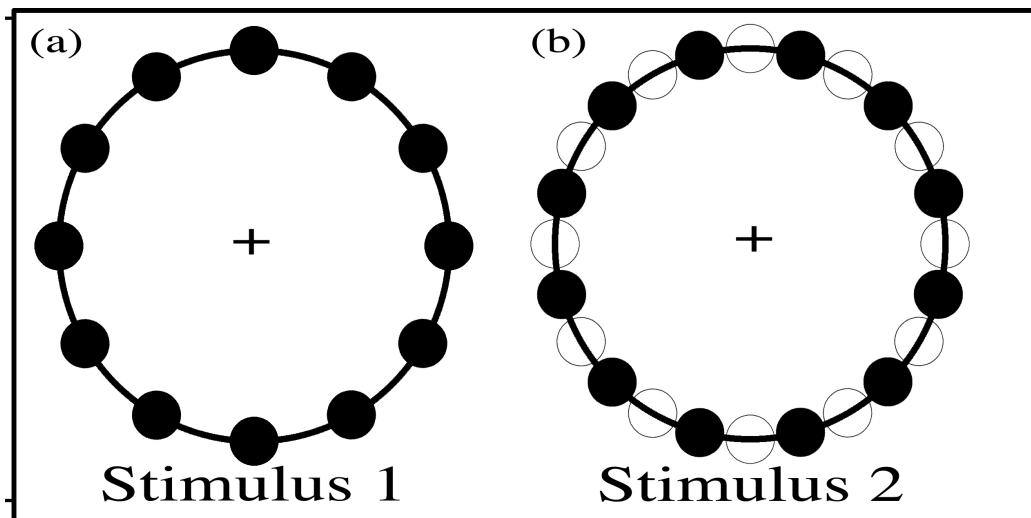
Figure 2.2: The visual stimuli used for this experiment. These were always presented as a pair with the stimulus on the left presented first. The outlines on the right are present to show where the first stimulus previously was seen. This pair of stimuli is symmetrical across many axes and the transition from the first stimulus to the second can cause a perception of rotation clockwise or counterclockwise.

ual radii and their distance from the screen's center), was selected to qualitatively maximize this effect. The initial circles are seen either as moving clockwise or counterclockwise around the screen. Minimal pretesting was sufficient to demonstrate that these two percepts were under the control of a practiced viewer and exhibited the same qualitative characteristics of standard static bistable stimuli - that is, the percept of motion did not interfere with the perception of reversals and stabilities.

2.4   Procedure

Potential participants responded to a blurb in Reed College's Student-Body Info, a weekly newsletter from and to the student body, flyers around campus, and a Facebook event which contained the same text as the blurb and flyer. Potential participants were screened before arrival at the lab according to the inclusion and exclusion criteria. Participants were informed of all study details before consenting, and were free to ask questions of the researcher throughout, as participant knowledge of the purpose or methods of research would not have interfered with results.

Some misconceptions can exist in the general population about neuroimaging techniques and for this reason, the researcher talked to participants when they arrived in the lab, and throughout the administration of the experiment to build rapport and make the participants feel more comfortable.

2.4.1 Pretests

Initial tests of visual acuity using the Snellen eye chart, and of subjects' tonal orientation were conducted before subjects were fitted with the EEG cap. The test of tonal orientation was administered in solitude in the same soundproof chamber and with the same audio apparatus as the remainder of the experiment. The test consisted of 240 Shepard tone pairs, (20 starting on each of the 12 pitches of the Western chromatic scale). Each pair of pitches fell six semitones apart to conform to the design of the tritone paradox, and each tone lasted 300 ms for a total stimulus duration of 600 ms. After each tone pair, participants responded with key presses to indicate whether they had heard an ascending or descending movement in the pitch. If they heard some ambiguous combination they were instructed to select the stronger percept.

Accuracy at this stage was extremely important for overall success, so ten minutes or more were sometimes taken to ensure good performance later on.

The result of this pretest was used to select the pitch class of the auditory stimulus to be played later in the following way. When all responses had been registered, the experimenter charted and graphed how many of these were heard descending, and was easily able to select the tone pair which had closest to half (10) responses heard descending. There were often two (or sometimes more) of these [28], which were considered interchangeable since there was no theoretical reason to preference one ambiguous tone over another. In rare cases when participants could not manipulate the first chosen stimulus effectively, one of these other, equally ambiguous tones were presented instead.

After this pretest, subjects were introduced to both auditory, and visual stimuli to be used in the experiment, and given several minutes to practice with each. Participants were informed at this stage of the bistable nature of both stimuli, and were encouraged to try to control their percept. When subjects reported that they could confidently control both stimuli, the experimenter instructed them to try to make the percept switch on as many trials as they tried to make it sustain itself.

2.4.2 EEG procedures

Participants were fitted with an electrode cap (similar to a swimming cap with 96 electrodes sewn into it) and with one free electrode placed on the face to measure eye blinks. Electrode impedances were kept under $5k\Omega$. In order to achieve this and ensure a strong connection between the electrodes and the skin, it was necessary to scrub the skin firmly with alcohol and water. A small amount of electrode gel was applied between each electrode and the scalp. This saline-based gel facilitates the detection of the small electrical signals produced by the participants' brains. To further improve the signal quality in the recordings the wooden part of a Q-tip was introduced through a hole in the electrode to push hair aside and to remove any air pockets to secure good contact between the scalp, the gel, and the electrode. It is important to note that this is a non-invasive technique, which allowed recording of the electrical activity generated by the participant's brain. No electricity was ever

sent to a participant's scalp, and the electrodes only record scalp electrical activity passively.

Participants were led to a recording room where they sat in a chair facing a computer screen and speakers. During the experiment, in order to reduce muscle electrical activity that interferes with the recording of brain activity, participants were asked to refrain from eye, head, neck, and body movements and eye blinking as much as possible. Because of these requirements participants were provided pre-arranged and/or participant-requested rest breaks throughout the experiment. Participants' bodies were centered on the screen, and the chair they sat in was adjusted such that their eyes were 75 cm from the screen throughout.

They listened and responded to auditory stimuli via a pair of speakers and viewed instructions and visual stimuli presented on a screen. All stimuli were instructional text or bistable sounds or images. Participants were asked to respond via button-presses to the second stimulus of each pair. The experiment consisted of between 1200 and 1800 trials for each participant, with each trial averaging 1.2s overall. Subjects who received more trials were those who had more difficulty reversing or sustaining the stimulus so that enough trials would fall in each category for statistically viable results. The judgment of whether to extend the experiment was made *in medias res* by the experimenter based on observations of participant responses in a realtime computer display.

Trials were broken down into three segments: a stimulus presentation for 300 ms followed by a second stimulus for 300 ms, and a brief inter-trial interval of 500-700 ms. The reason for this variable interval was to prevent any systematic tail end of a previous ERP from being collected in the following trial. Trials were divided into blocks of 300 which in turn were divided into sub-blocks of 60 so that participants were able to take short breaks every minute, and longer breaks every few minutes. The study took approximately 45 minutes to run in each case in addition to the time required to run pretests, prepare the scalp and place the electrodes for a total of approximately 2.5 hours.

At the end of the experiment, participants had the option to wash the electrode gel off of their faces and out of their hair.

2.4.3   Data Collection

The Caps used in this study were 96 channel Herrsching DE-82211 "Easycap" caps, manufactured for professional EEG data collection. The stimuli were presented on a Planar SA2311w 23" computer monitor, which could display 1920 x 1200 pixel resolution. The participant's head was located .75 m from the screen, causing the monitor screen to fill $57.3°$. For the visual trials, the stimuli filled $9.8°$. The computer monitor was located in a soundproof room construdted by *Industrial Acoustics Company Inc.* Auditory stimuli were presented on THX Logitech stereo speakers located on either side of the computer monitor. The same sound emanated from both speakers.

The electrodes on the EEG cap were connected to three 32 channel BrainVision "Professional BrainAmp" amplifiers. The gain is about 750 for BrainAmp Standards with a fix resolution of 0.1 μV and a fix range of $\pm 3,2768$ mV and the input noise

of $< 2uppV$. The amplifier also converted the signal from analog to digital at a rate of 500 Hz. The recording software imposed a bandpass filter, (low = .1 Hz, high = 150 Hz) on the incoming data because signals outside this range were unlikely to have originated in the brain and therefore would read only as noise, and displayed it in realtime on a monitor outside the recording booth so that the researcher could make changes and fix problematic channels throughout the experiment. The recording software also received codes from the stimulus machine corresponding to each stimulus (whether visual or auditory) with each stimulus of a pair having a different code, as well as from participant key-presses. This data was integrated with the incoming electrical signals, and the entire dataset was saved as a trinity of files to be read by Analyzer software (Brain Products, Germany).

2.5   Data Analysis

2.5.1   Artifact Rejection

After initial import into Analyzer, the complete dataset for each subject was segmented into eight segment types, which were grouped together. Segments were created identically for visual and auditory. In each modality, perceptual reversals and stabilities were coded. The former was accomplished by finding stimuli with different responses within a 1200 ms time window before and after, and the latter by finding stimuli with the same response before and after within this window. The result was that datasets were now available in both modalities for each percept type (percept A and percept B), and for each behavior type (reversal or stability). Reversals and stabilities were coded via a boolean operation which looked 1200 ms before and after each stimulus onset. If participants had responded with the same button press in each of these time-windows, then the trial was considered a stability, but if the button presses differed, then the trial was coded as a reversal. Four comparisons were planned between these eight conditions - percept A versus percept B and reversal versus stability for each modality.

The channels to the left and right of the eyes, LHEOG (84) and RHEOG (72) were combined into a single HEOG channel by referencing one to the other

This procedure served to make any eye movements more obvious. If any channels were perceived to be prohibitively noisy, these were also replaced at this stage with a pool of surrounding channels.

All channels were re-referenced to the average of the mastoid channels (74, 82).

Artifacts were removed in three similar steps, each using slightly different criteria and different channels. First, eye movements were sought and removed using the newly created HEOG channel. This was done with a measure of absolute difference in the height of the waveform over a 50 ms interval, with a sensitivity of 50 μV. Next, vertical eye movements and blinks were treated in the same way, this time using channels above and below the eye (71, 96). In this case, criteria were a min-max difference in excess of 100 μV within a 300 ms time window. Finally, both a max-min measure, and an amplitude measure were applied to all other channels to discover any

remaining noisy segments. In this case, a maximal allowed absolute difference of 125 μV over a 300 ms time window was employed, and maximal absolute amplitude of 150 μV was also used. The ranges for these criteria differed between subjects because different channels were noisy on different heads, and because some individuals blinked or glanced to the side more often or more forcefully. The goal was to remove nearly all artifact-ridden segments without removing any clean segments, and this *sweet spot* was different for different subjects. What is reported here is a set of exemplary values and not a standard set used for all subjects.

After artifact removal, all signals were passed through a low pass filter (upper bound set at 30 Hz.), combined with a 60 Hz. notch filter. The former removed much of the high-frequency noise which was unlikely to correspond to any meaningful signal from the brain, and the latter significantly reduced noise due to ambient electrical sources since 60 Hz. is the frequency of the AC house current from the American power grid. These two filters are compatible because the 30 Hz. filter decreases the amplitudes of progressively higher frequencies in a gradual manner, so some 60 Hz. noise was still present.

All segments were then averaged together, and baseline corrected against the 100 ms immediately preceding the segment.

When the above procedure had been completed for all eight segment types for each subject, all subjects' data were combined into a grand average.

2.5.2   Selecting Segments for Analysis

Once grand averages of each of the eight conditions had been created, the four comparisons listed above were applied, and difference waves were constructed for the comparisons. Each difference wave was visually inspected in reference to prior literature regarding bistability, and temporal regions of interest (ROIs) were identified as any area in which this waveform exceeded $.5\mu V$. Scalp topographies were also generated from these difference waves in multiple frames, usually corresponding to 50 ms each. These were also inspected spatially and temporally for ROIs. At times it was difficult to tell whether a given ROI should be analyzed as one component or as two, and in these cases, the difference wave was used to determine the location of peaks and troughs.

2.5.3   Statistical Analysis Procedures

For each ROI, a dataset was exported from Analyzer containing mean amplitude measurements in each condition for each subject. The dataset was constrained to the specific time-window of interest in each case, but included all 96 electrodes. This dataset was imported into StatSoft Statistica 10, and RM-ANOVAS were used to determine the statistical significance of each ROI.

2.6 Procedural Changes

While all numerical choices reported remained constant throughout the experiment, several procedural elements underwent small changes during testing. Of these, the most important was the method of administering the tonal orientation pretest. While the pretest is largely an unambiguous measure of tonal orientation and based on the test's similarity to prior research and the similarity of its results, it was an accurate measure as well, the experimenter began to suspect that the most ambiguous tone was not always the easiest one to manipulate perceptually. For this reason, strict adherence to the results of the pretest was relaxed in later testing such that participants who reported difficulty would be offered a second and sometimes even a third tone pair in search of the most bistable stimulus possible. In all events, the most bistable stimulus was sought, regardless of its relation to the subject's tonal orientation, although in no subjects did this priority result in a wild discrepancy.

This procedural change was employed because the goal of this investigation was not to confirm or replicate or even expand upon the literature on the tritone paradox, but rather to test if such a stimulus could be used in an EEG paradigm and to use it to probe the auditory system.

Chapter 3

Results

3.1 Results of Pretest

The pretest of peak pitch class successfully performed its intended function of determining candidates for ambiguous stimuli. After taking the pretest, all subjects were able to find and learn to control a bistable stimulus from among the twelve possibilities. As expected from Deutsch's research, no trends were apparent in the set of peak pitch classes from the 14 subjects, or in the positions of the most ambiguous tones. Indeed, at times, subjects showed diametrically opposite responses to the pretest, resulting in peak pitch classes differentiated by 5-6 semitones. This opposition is exemplified by the subjects whose data are pictured in Fig. 3.1 (a) and (b). Notably however, all subjects showed the same general contour - a region of high percentage of descending percepts, and a region of low percentage. When peak pitch class was taken as a constant factor, and all subjects' data sets were aligned by this pitch class and then percentages of descending responses were averaged together, a smooth and regular trend was observed, again as was found by Deutsch. This trend is displayed in Fig. 3.1 (c).
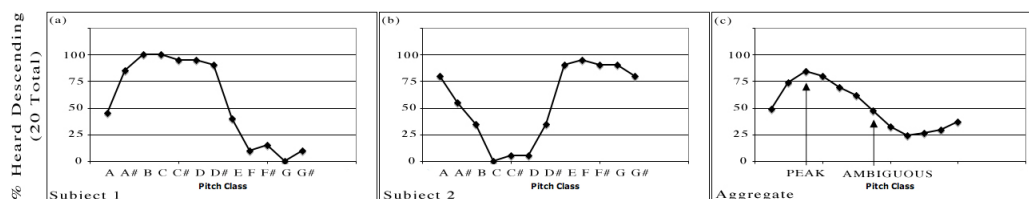
Figure 3.1: The results of the pretest in this experiment were very much in line with the prior results found by Diana Deutsch, and this is to be expected due to the similarity between this test and Deutsch's paradigm for finding peak pitch class. The first two panels (a) and (b) are two individual subjects, demonstrating the strength of individual differences in responses to the paradox. Panel (c) is an aggregate of all subjects with peak pitch classes aligned at the third position. This panel demonstrates that the high and low regions are very clear across subjects, as is the ambiguous intermediate point. "Pitch Class" on the horizontal axis refers to the initial tone.

3.2   Behavioral Data

The goal of this experiment was not necessarily to achieve a perfectly equal number of trials in each condition in each case, but a decent balance was sought to allow fair comparisons between ERPs of reversal and stability. This goal was well met, although small differences did occur between conditions. While most participants were tested for a standard 600 trials for each modality (1200 total), some were administered extra trials for a variety of reasons (difficulty of task, incorrect responding, etc.). Normalized components were thus employed instead of raw numbers of trials. Tbl. 3.1 reports the mean number of responses in each condition and the standard deviations as percentages.

Table 3.1: Responses in each condition, averaged across subjects, and reported as a percentage. The first four entries are components analyzed in this investigation, while the final four are included to demonstrate that subjects' percepts were not weighted more toward one or the other.

Condition	Mean	Standard Deviation
Auditory Reversals	44.6%	7.2%
Auditory Stabilities	55.4%	7.2%
Visual Reversals	40.6%	9.1%
Visual Stabilities	59.4%	9.1%
Visual CCW	50.8%	8.0%
Visual CW	49.2%	8.0%
Auditory CCW	50.7%	3.3%
Auditory CW	49.3%	3.3%

Reaction time was another quantity which was measured in this experiment as a natural consequence of the design. Between the four conditions, there was not much difference, especially compared to the immense standard deviation in each condition. The data in Tbl. 3.2 record this, and it should be noted that the large standard deviations are not due to a small number of trials. The minimum number of trials per condition was greater than 3000, so the population of 14 college students tested in this investigation did indeed vary to great extent the timing of their responses to the second stimulus. Notably, the reaction time was not noticeably different for visual and auditory reversals, which begins to opposes the published suggestion of Pressnitzer et. al. in 2006 that auditory manipulations were more difficult and took longer [80].

Table 3.2: Reaction times for each condition. Reaction times were averaged across all participants' responses within a condition, but responses of greater than 1100 ms were excluded because these would possibly be taking place in the time-window of the following trial. Responses of less than 250 ms were also excluded because these were considered artifacts, due to such extreme deviation from prior reaction time estimates.

Condition	Mean (ms)	Standard Deviation (ms)
Auditory Reversals	414.86	142.60
Auditory Stabilities	399.57	130.72
Visual Reversals	322.21	73.61
Visual Stabilities	318.91	78.42

3.3   ERP Components of Interest

Previous studies of visual RN and LPC were used to guide identification of components for analysis. For additional components, visual inspection of the superimposed waveforms for the two conditions (reversal and stability) for each modality, and inspection of the difference waves were used to determine regions of interest (ROIs). Time windows (40-60ms) and electrode locations that best characterized the spatiotemporal properties of each ROI were then selected.

Overall, seven ROIs were isolated, corresponding to the strongest signals on the scalp during any time window. Of these, three were of especial interest as potentially matching components from prior literature. A visual RN was observed which seemed to match temporally and locally the RN which has been previously described. Components which resembled LPCs for both visual and auditory modalities were also observed, although these differed slightly in timing and location with the auditory component taking place some ms. later and more frontally. Four further components were observed, two visual positivities and two auditory negativities, all located roughly at the vertex of the head, and these do not match the timing or location of previously observed components. These seven components are indicated in both Fig. 3.3 and Fig. 3.4. Statistics and further information for these components can be viewed in Tbls. 3.3 and 3.4.

The distinctions between the three positive visual components and between the two negative auditory components were made via examination not only of the ERPs themselves, but also the difference wave, which showed clear peaks and troughs between the components proposed here (Fig. 3.2).

Mean amplitudes for each component were analyzed for statistical significance with a mathematical software StatSoft Statistica 10. Each component was analyzed via a separate repeated measures analysis of variance (RM-ANOVA), to determine whether condition (reversal or stability in each case) had any effect on electrical
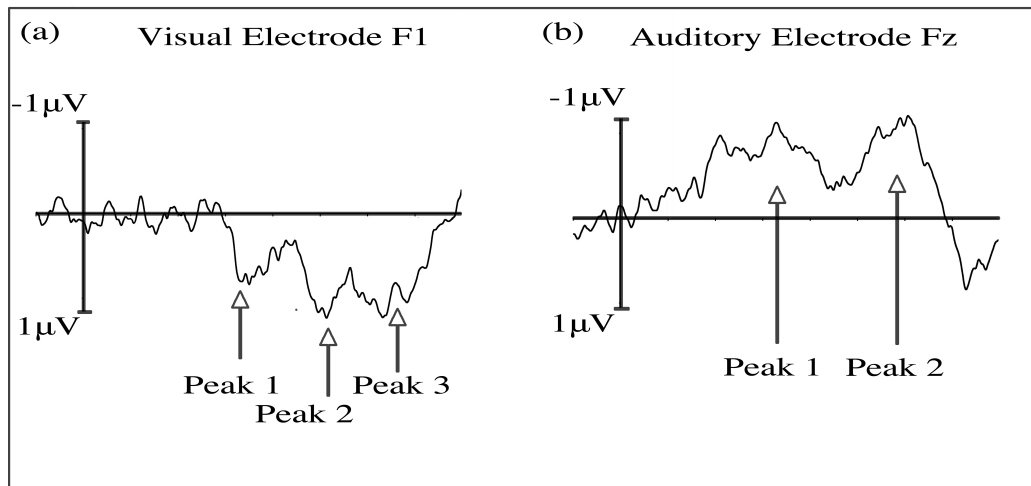
Figure 3.2: In cases where visual inspection of the ERPs was difficult, or was insufficient to determine the number of components, the difference wave between conditions was used as well. The panels pictured here show the difference wave, formed by subtracting the amplitude of the stable ERP from the amplitude of the reversal ERP. Panel (a) now clearly shows the three positive components in the visual modality, while panel (b) shows the two negative components in the auditory modality.

activation. Statistically significant main effects of condition were found for five components, and near significant main effects were found for two further components. These results are summarized in Tbl. 3.3. The means on which these ANOVAs were based are reported in Tbl. 3.4

3.3.1 First Visual Positivity

A positivity was observed in the visual data at around 275 ms, and peaking 80 ms later at 355 ms. This component was localized at the central scalp, with the reversal condition appearing more positive than the stable condition. An RM-ANOVA was performed to determine whether reversal was more positive in terms of mean amplitude than stable over the time-interval from 310 - 370 ms. There was a statistically significant main effect of condition, $(F(1, 13) = 7.74, p < .02)$ over the time interval. The average amplitude of reversals was 0.45 μV (SD = 0.838), and the average amplitude of stabilities was -0.267 μV (SD = 0.705). This component is visible in Fig. 3.3 (a), and the values reported here are summarized in Tbls. 3.3 (ANOVA) and 3.4 (means).

3.3.2 Second Visual Positivity

A second positivity was observed in the visual data at around 440 ms (140 ms after the second stimulus), peaking at 505 ms. This component was also localized at the vertex, with the reversal condition appearing more positive than the stable condition.
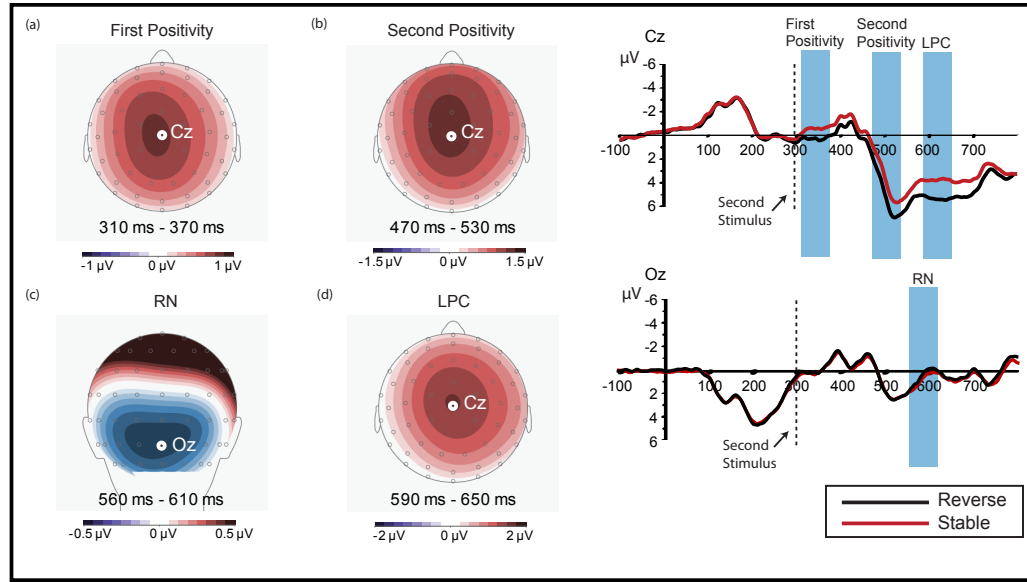
Figure 3.3: Four visual ROIs were investigated, represented by the head views at left. A first and second visual positivity (a) and (b), a visual RN (c) and a visual LPC (d). The three positive components are statistically significant, while the RN is near significant. The time windows from which these head views are drawn are displayed at right, situated within the complete waveform at a single electrode. Note that all views were from the top with the nose upward, except for the RN, which is shown from the back.

An RM-ANOVA was performed to determine whether reversal was more positive in terms of mean amplitude than stable over the time-interval from 470 - 530 ms. There was a statistically significant main effect of condition, $(F(1,13) = 8.50, p < .02)$ over the time interval. The average amplitude of reversals was 4.42 μV (SD = 0.845), and the average amplitude of stabilities was 3.286 μV (SD = 0.682). This component is reported in Fig. 3.3 (b), and the numbers reported here are summarized in Tbls. 3.3 (ANOVA) and 3.4 (means).

3.3.3   Visual RN

A negative component was observed in the visual data at around 560 ms (260 ms after the second stimulus was shown), peaking 35 ms later at 595 ms. This component was localized over the central occipital scalp, with the reversal condition appearing more negative than the stable condition. An RM-ANOVA was performed to determine whether reversal was indeed more negative in terms of mean amplitude than stable over the time-interval from 560 - 610 ms. There was a statistically near-significant main effect of condition, $(F(1,13) = 3.070352, p < .1)$ over the time interval. The average amplitude of reversals was 0.198 μV (SD = 0.395), and the average amplitude of stabilities was 0.567 μV (SD = 0.301). This component is reported in Fig. 3.3 (c), and the numbers reported here are summarized in Tbls. 3.3 (ANOVA)

and 3.4 (means). This component is included here, despite the sub-significant statistical analysis because it closely matches what is expected of the RN component from prior static bistable stimuli.

3.3.4   Visual LPC

A third visual positivity was observed, starting at around 580 ms (280 ms after the second stimulus was shown), peaking at 630 ms. This component was also localized at the vertex, with the reversal condition appearing more positive than the stable condition. The rationale for analyzing these three positive visual components separately arose from visual analysis of the difference wave at electrodes at the vertex. In these electrodes, although the voltage remains positive throughout the 800+ ms of the whole trial, three separate peaks are differentiable. While this is in itself a sufficient reason for analyzing these separately, a further point is that for this positivity to reflect a single long-term component, it would need to last for 500+ ms, which is unlikely at best given the nature of the stimuli (visual geometries), and the onset time of the first positivity (before stim 2 onset, ¡ 300 ms). An RM-ANOVA was performed to determine for this third positivity whether reversal was more positive in terms of mean amplitude than stable over the time-interval from 590-650 ms. There was a statistically significant main effect of condition, ($F(1, 13) = 12.156, p < .005$) over the time interval. The average amplitude of reversals was 5.334 μV (SD = 0.779), and the average amplitude of stabilities was μV 4.138 (SD = 0.612). This component is reported in Fig. 3.3 (d), and the numbers reported here are summarized in Tbls. 3.3 (ANOVA) and 3.4 (means).

3.3.5   First Auditory Negativity

A negativity was observed in the auditory data, peaking at 265 ms (even before the second stimulus onset). This component was localized at the vertex, with the reversal condition appearing strongly more negative than the stable condition. An RM-ANOVA was performed to determine whether reversal was more negative in terms of mean amplitude than stable over the time-interval from 250 - 290 ms. There was a statistically significant main effect of condition, ($F(1, 13) = 7.608, p < .02$) over the time interval. The average amplitude of reversals was -3.893 μV (SD = 0.458), and the average amplitude of stabilities was -2.905 μV (SD = 0.478). This component is visible in Fig. 3.4 (a), and the numbers reported here are summarized in Tbls. 3.3 (ANOVA) and 3.4 (means).

3.3.6   Auditory RN

A second negativity was observed in the auditory data, peaking at 545 ms (245 ms after the second stimulus). This component was found on top of the head, with a near significant lateralization to the left, with the reversal condition appearing strongly more negative than the stable condition. An RM-ANOVA was performed to
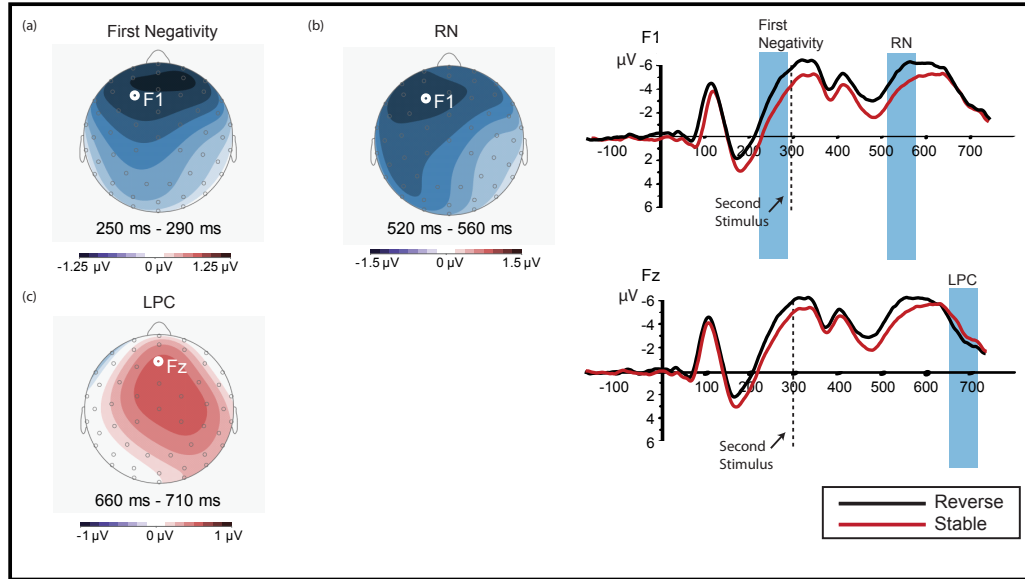
Figure 3.4: Three Auditory ROIs were investigated, represented by the top views of the scalp at left, showing initial first auditory negativity (a), an auditory RN (b), and an auditory LPC (c). The two negative components are statistically significant, while the LPC is not, but matches the time, sign and location of a component reported previously. The time windows from which these head views are drawn are displayed at right, situated within the complete waveform at a single electrode which clearly showed the trend.

determine whether reversal was more negative in terms of mean amplitude than stable over the time-interval from 520 - 560 ms. There was a statistically significant main effect of condition, $(F(1, 13) = 8.377, p < .02)$ over the time interval. The average amplitude of reversals was -4.914 μV (SD = 0.627), and the average amplitude of stabilities was -3.843 μV (SD = 0.543). There was a near-significant condition x hemisphere interaction, $(F(1, 13) = 2.44, p = 0.142)$. Although this interaction was nonsignificant, it is included here as a potential direction for future investigation. This component is visible in Fig. 3.4 (b), and the numbers reported here are summarized in Tbls. 3.3 (ANOVA) and 3.4 (means).

3.3.7 Auditory LPC

A negativity was observed in the auditory data, peaking at 670 ms (370 ms after the second stimulus onset). This component was again locted at the vertex, with the reversal condition appearing strongly more positive than the stable condition. An RM-ANOVA was performed to determine whether reversal was more negative in terms of mean amplitude than stable over the time-interval from 660 - 710 ms. There was a statistically significant main effect of condition, $(F(1, 13) = 7.608, p < .02)$ over the time interval. The average amplitude of reversals was -3.893 μV (SD = 0.458), and the average amplitude of stabilities was -2.905 μV (SD = 0.478). This component

is visible in Fig. 3.4 (c), and the numbers reported here are summarized in Tbls. 3.3 (ANOVA) and 3.4 (means).

Table 3.3: Seven ROIs were investigated, four visual and three auditory. Peak times, and time windows used for analysis are reported here, as are F and p-value results form separate RM-ANOVAS performed for each component.

Component	Peak Time	Time Window	F	p
Visual Positivity 1	355	310 - 370	8.50	0.012031*
Visual Positivity 2	505	470 - 530	7.74	0.015551*
Visual RN	595	560 - 610	3.07	0.103269
Visual LPC	630	590 - 650	12.16	0.004016*
Auditory Negativity 1	325	250 - 290	7.61	0.01628*
Auditory RN	605	520 - 560	8.38	0.0125*
Auditory LPC	730	660 - 710	1.24	0.29

Table 3.4: Means and standard deviation on which the seven RM-ANOVAs were based. These means are mean amplitude of the component over the time-window chosen for analysis. These time-windows are reported in Tbl. 3.3. Mean 1 here is reversals and mean 2 is stabilities. All entries are reported in μV.

Component	Mean 1 (Reverse)	SD	Mean 2 (Stable)	SD
Visual Positivity 1	0.45	0.838	-0.267	0.705
Visual Positivity 2	4.42	0.845	3.286	0.682
Visual RN	0.198	0.395	0.567	0.301
Visual LPC	5.334	0.779	4.138	0.612
Auditory Negativity 1	-3.893	0.458	-2.905	0.478
Auditory RN	-4.914	0.627	-3.843	0.543
Auditory LPC	-2.658	0.462	-3.118	0.495

Chapter 4

Discussion

4.1 Summary of Results

For both the bistable apparent motion stimulus and the tritone paradox auditory stimulus, ERP amplitudes differed on trials in which subjects reported perceptual reversals compared to perceptual stability. While some of the amplitude differences were consistent with previous reports, others were novel and were made possible by the particular paired stimulus presentation employed here. Seven components (Secs. 3.3.1-3.3.7) were observed, four visual and three auditory.

In the visual domain, a nearly significant RN and a significant LPC were observed at very similar times and positions as in prior bistability studies. Two further positive components were observed at the vertex earlier than the LPC, and these are proposed to have some function related to intending or anticipating a reversal.

For audition, an LPC was also observed, although it did not reach significance, and was located more frontally than that for vision. Beyond this, two to four (two are reported here) large negative components were observed over the frontocentral scalp, slightly left-lateralized in the period from 200 - 700 ms. One of these negativities may reflect an auditory analog of the RN.

4.1.1 Visual RN

A bilateral negativity was observed at occipital and parietal scalp sites, beginning about 280 ms after the second stimulus, and this activity matches closely the characteristics of the visual RN. This component is thought to be related to the event of a perceptual switch, although the exact relationship is unknown. While the visual RN in this study did not quite reach significance, it is reported here because the timing, location and sign of this component match prior reports of reversal negativity. Thus the presence of this component confirms that these visual stimuli function as bistable in the brain as well as behaviorally, and matches previous studies with the SAM stimuli which also found this component.

4.1.2   Visual LPC

A positive component at the vertex matched the location, sign and timing of the LPC. Together, this component and the visual RN act as a replication of prior literature, albeit using a different stimulus. The LPC is associated with noticing that a switch has occurred, regardless of whether it was created endogenously or exogenously, and although the RN remains in both cases as well, it shifts earlier when generated endogenously.

4.1.3   Visual Positivities

Two positive components were found in the visual modality over the vertex at the time of the second stimulus's presentation, far too early to be properly considered as responses to the second stimulus. These components may not be stimulus related at all and instead may be indicative of the brain's focus on causing a reversal, or on anticipating an approaching reversal. This proposition is supported by the location and timing of these components which do not appear to come from the visual cortex, but rather from frontal areas, and which seem almost to precede the bistable stimulus. Due to the inverse problem, it is impossible to make any spatial claim with complete confidence, but *some* spatial resolution does exist, and it is enough to indicate whether sources are located frontally or occipitally. In this case, the positivity was localized to the vertex.

Previously, a similar central positivity has been observed in intentional switching relative to incidental switching of a static bistable stimulus [76]. This is certainly an area which deserves further investigation, as to date it has not previously been examined directly, (nor was it here), and as it is still unclear whether this positivity is a single, extended component or two overlapping components which further overlap with the LPC.

4.1.4   Auditory LPC

The auditory LPC, which did not reach the cutoff for significance, is included in this discussion for the same reason as its visual double. Even though this component was small in amplitude, it matches prior literature for the timing, sign and location of the LPC, and is therefore included here as evidence of potential convergence across modalities.

Interestingly, this LPC is more frontal than for vision, although it remains on the midline. The cause of this change is unknown and requires further exploration but it may be due to the more anterior location of the auditory cortices compared to the visual cortices.

That this component did not reach significance is somewhat troubling. One possible reason for this is that this component is riding on the heels of much larger negative components, so it is possible that in relation to the general state of this temporal region, this component is actually much stronger, and is simply being masked by the negative components. Another possibility is that the time taken to reverse is

longer for auditory stimuli as reported previously [80] and also noted from qualitative participant reports in this study, with the result that reversals did not occur at a particular time for all subjects and all trials, or that reversals did not even occur at a well-defined time. If these explanations are the case, the smaller amplitude for the LPC is expected, and the peak might be more spread out.

4.1.5 Auditory Negativities

The auditory difference waves are characterized by one or more strong negative components located at the central scalp, and lateralized slightly to the left. Since this modality and design have not been explored previously, it is not clear what should be made of these components. One of these negativities may be the auditory analogs of the visual RN, while the other may be an indicator of the intention to switch. The latter component would be easy to test with subjects in different experimental conditions receiving different instructions: to switch intentionally as often as possible, or to allow reversals to take place naturally. The first negative component would be predicted to disappear in the condition with no instructions but to remain in those participants who attempted to exert control. The latter could be tested by constructing an un-ambiguous version of the tritone paradox (an easy goal by playing with the relative amplitudes of the component frequencies) and inducing reversals exogenously. If the component is indeed an auditory analog of the visual RN, it should still be apparent when the reversals are induced via the senses rather than naturally generated in the brain. Regardless of their interpretations, these components are very strong and deserve replication and further investigation.

4.2 Limitations

4.2.1 Ambiguity

Two important notes on the subject of bistable percepts relate to ambiguity, which may have caused some differences between subjects in hearing the tritone paradox. The bistable percept will be perceived one way, then the other, but never both at once, whereas an ambiguous stimulus can give rise to both percepts simultaneously. This distinction is of great importance to experimenters because of these two types of stimuli, only bistable ones can cause perceptual reversals. The tritone paradox exists in a strange class because although all the sensory information is present for it to be perceived ambiguously, participants report clear rising or falling perceptions. Two subjects, however, did report the ability to hear the motions of multiple octaves of the tones which made up the paradox. These participants were instructed to focus on the louder pitches, but it is possible that these subjects were indeed hearing opposing motions simultaneously.

The majority of bistable percepts are in fact multistable in that there is usually a third version of the percept which is akin to the raw data input to the perceptive system. Those versions of the percept which contain coherent information are more

attractive in some sense to the attention of the viewer, and it is sometimes difficult to break away from these to see the raw form again. To look closely at the lines of the Necker cube in Fig. 1.3 (a) is to be able to see a flat, unformed set of diagonals, verticals and horizontals which are the raw data whereas the two stable cubes which unify the disparate linear information are much easier to see.[1]

4.2.2   Different Stimuli

The purpose of using bistable stimuli is, of course, to minimize differences between experimental conditions, and while this goal was achieved, the use of the Tritone Paradox in this way introduced a new type of confound, namely, that each subject was hearing a different set of frequencies based on the result of their pretest. This is a valid point, and although it appears not to have nullified the results in this case, it must be noted and addressed.

One obvious solution is to pretest a large enough population that a sufficient number of subjects share a peak pitch class, *i.e.* the same stimulus can be used for all of them. Another solution is to change the paradigm somewhat, choosing a specific key first and training subjects to hear and control the tritone paradox until they can be tested regardless of their innate peak pitch class. Either of these methods will work, but neither is ultimately necessary, nor indeed desirable. With respect to desirability, in the case of the first solution, many more subjects (12 times as many) require testing, whereas for the second solution it becomes impossible to test naive or nearly naive subjects. Furthermore, in this case, the added factor of voluntary control is built into the experiment and it is not at all certain that this factor is itself constant across all subjects.

As to necessity, before searching for a solution however, one must always ask whether the issue at hand is indeed causing problems with the data at all. In this case, the answer may be that it is not. While different subjects certainly heard different sounds, and this created differences between subjects, all subjects were hearing the *most bistable* tone for them, meaning that instead of keeping the specific tones constant, ambiguity was kept constant (maximized).

Another way to think of this is that instead of keeping pitch height at a constant value, peak pitch class was held constant across subjects. If the same tones were used for all subjects, pitch height would be constant, but peak pitch class would then be randomly distributed with respect to the stimuli and this factor would be much more likely to skew the results.

This problem has been faced previously [47], and was dealt with in the same way. These researchers argued that maximizing ambiguity is preferable to homogenizing

---

[1]This is a similar effect to that seen in subjects' responses to binarized images (that is, images reduced strictly to black and white) which can become meaningless blotches of the two colors. When subjects are shown the binarized version they see nothing, but if they are then shown the image without binarization to show them what is present, and again shown the binarized version, the meaningful lines are easily apparent [32, 94]. As a simpler figure, the Necker cube is easier to perceive as a set of unrelated features, but the dominance of a percept which binds the features together is present in either case.

the stimuli. Finally, this confound exists between subjects, rather than within subjects, and thus the two conditions being compared (reversals and stabilities) are still being tested with identical stimuli in each subject.

4.2.3   Locus of Bistability

An important concern when using auditory stimuli constructed in this manner is where the bistability occurs. Most listeners qualitatively describe the second tone out of a tritone pair as the one which sometimes sounds higher and sometimes lower than the first, but no quantitative test exists to confirm these reports. Furthermore, a handful of listeners report hearing multiple distinct frequencies in both the first and the second Shepard tones as in Fig. 4.1 (c). This is a worrisome comment because if multiple sounds are being heard within the first tone, then it is possible the bistability is taking place in the first sound, rather than the second, and listeners are hearing first a descent from a high initial frequency to a neutral second frequency, and then an ascent from a low initial frequency to the same neutral final frequency (Fig. 4.1 (b)).

One possible solution is to change the initial tone from a Shepard tone (Section 1.2.3) to a pure tone, making its frequency completely unambiguous to the listener. This would force the bistability into the second tone, but it is unclear how it would affect the completeness of the illusion, since most people can hear the difference between a pure tone and a tone constructed from many frequencies and this might provide too much of a reference point and thus render the second tone unambiguous as in Fig. 4.1 (a).

The answer to this question certainly affects the interpretation of the data because all components have been discussed here in relation to the second stimulus. It is unlikely that this was a widespread problem in the present investigation, nor does Diana Deutsch mention anything like this anywhere in her published literature on the paradox, but it should certainly be taken into account when these stimuli are used again.

One possible solution is to change the initial tone from a Shepard tone (Section 1.2.3) to a pure tone, making its frequency completely unambiguous to the listener. This would force the bistability into the second tone, but it is unclear how it would affect the completeness of the illusion, since most people can hear the difference between a pure tone and a tone constructed from many frequencies and this might provide too much of a reference point and thus render the second tone unambiguous as in Fig. 4.1 (a).

4.2.4   Musical Training and Spectral Envelopes

Following Diana Deutsch's initial investigation of the tritone paradox, she endeavored to contextualize her results by expanding her population of study and the mechanical diversity of her stimuli. The first of these was accomplished by changing from a musically trained population (defined as having two or more years of musical training), to one which was not constrained by training. This manipulation served to increase the number of participants tested overall, to replicate her results from her first experiment, and showed that not only was the tritone paradox easy to hear and respond to for most people but that nothing about the procedure (ease of responding, number of stimuli heard ascending versus descending, or peak pitch class) was dependent on musical training.

Peak pitch class is related to vocal range and native region, but is close to randomly
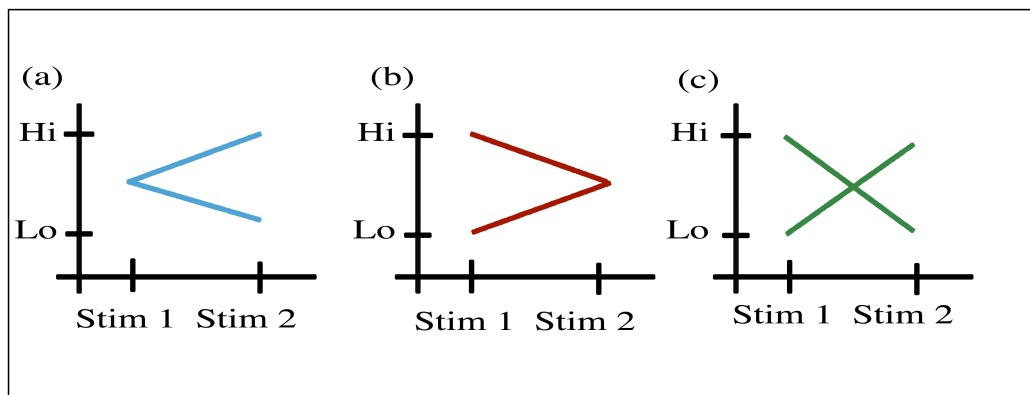
Figure 4.1: Panel (a) shows what the stimuli in this investigation were originally suggested to be doing, and what they would be forced to do if the change proposed in Sec. 4.2.3 were to be made. Panels (b) and (c) are other potential bistable arrangements for these stimuli which might have been heard and therefore affected the data. The bulk of the qualitative evidence from participant report, is against this, but scattered reports do cast doubt.

distributed in a college population which includes persons of many origins. This result is important because it means that all participants could be considered, rather than splitting the participant pool into those with musical training and those who are musically naive.

The second test, of the spectral envelope around the stimuli, achieved statistical results, though small in size. Shifting the spectral envelope which encompassed the tones used in her experiment up by 6, 12, and 18 semitones (a constant shift relative to the musical tones in the paradox, but not to the frequencies), Deutsch found that the contour of the response curve changed (peaks and troughs in the graph of "% heard descending" versus "pitch class" may broaden or sharpen), but that these changes were minimal and the maximal change in peak pitch class when the spectral envelope was altered was not more than about one semi-tone.

In creating the stimuli for my own investigation, previously created stimuli from Deutsch's website were transposed upward or downward by one or two half steps, so that from the original 4 tones, a full 12 could be used. Because of the way that Deutsch created her stimuli (such that they share a spectral envelope), this means that more than one spectral envelope was used in the present test. These envelopes were neither consistent with respect to an absolute frame (as Deutsch's were), nor with respect to the pitch class of the initial tone of the paradox.

The response to this issue is the same as that in Section 4.2.2, that while the spectral envelope may have a small effect on the shape of the distribution of descending responses, the target in the pretest was not peak pitch class as it was in Deutsch's investigation, but instead was maximal ambiguity and therefore bistability, which was achieved regardless of the spectral envelope.

4.2.5   Pretest of Peak Pitch Class

Although the pretest used in the present experiment was nearly a perfect replication of Deutsch's test, and acquired similar results along the dimensions she measured, it is not certain that for each subject the most ambiguous looking stimuli were indeed the most readily reversible. This was discovered when a small subset of subjects complained of the difficulty of the task (which was at the time programmed to utilize the most ambiguous stimulus based on the pretest). The experimenter chose the next most ambiguous stimulus, and at times even the third most ambiguous, and subjects often reported an increase in control over the perceptual reversal. This unexpected fact may be due in part to the issues raised in Section 4.2.4, or may be due to a placebo of some kind, or may simply indicate that this pretest does not result in the most bistable of a set of 12 stimuli. Whatever the reason, the experimenter in each case followed the advice of prior researchers with a similar pretest [47], and again maximized ambiguity, rather than strictly following the results of the pretest.

4.3   Implications

While SAM stimuli have been used in the past the present research is, to my knowledge, the most complete investigation of the phenomenon of apparent motion using an EEG paradigm. In this case, both the LPC and the RN common to static stimuli were observed, establishing dynamic visual stimuli as viable bistable images. This demonstration is an important change as many paradigms require some movement of stimuli in order to create an effect. Now the methods of perceptual analysis of bistable stimuli are also potentially available to combine with these methods. Some creativity will need to be used to combine established paradigms, but this can now confidently be done with the knowledge that the stimuli do indeed function bistably.

Auditory bistability is a young field, and the advent of an auditory stimulus which matches the criteria necessary for EEG is an important development. Further experimentation remains to be performed to test whether the two modalities under study here do indeed share mechanisms of perception.

Of surmounting importance for future research is that *two* modalities are now available for testing and so by comparison between matched stimuli in these two, judgments can potentially be made as to which components are sensory (and therefore will admit unique signatures from the two systems), which are perceptual (and therefore may or may not admit similar features) and which are top-down processes of attention, effort or expectation (and therefore are very likely to resemble one another).

With stimuli which are made of two parts, the first stimulus sets up and immerses the subject in the illusion, and gives reliable information as to when the second (ambiguous) stimulus will appear. This allows, for the first time, the subject keenly to anticipate the arrival of the second stimulus and make preparations for that arrival. Exploration of this pre-stimulus interval may prove distinctly fruitful, as it already has in the present investigation, and is only possible with stimuli which arrive in multiple parts. A beneficial addition to this paradigm would be a comparison condition in

which the second stimulus is presented alone and this ERP compared to that which resulted from bistable perception to tell which components were present due to the stimulus and which were due to the paradigm itself.

Potentials related to the intention to switch are an interesting area for future research, and have been found now in two studies, the present one, and in a study by Pitts et. al. [76]. Investigating the effects of intention and expectation (which can likely be independently manipulated) at the boundary between sensation and perception may prove a compelling and rewarding future research path.

4.4   Future Directions

While both aspects of the present investigation are new directions in the field of EEG research, and certainly bear replication before future directions are determined, some further questions do present themselves now.

First, this experiment was not designed with the pre-stimulus interval in mind, and therefore is only able to probe that time-region superficially. A further test maintaining many aspects of this design, but focusing on consistent instructions to participants and specifically perhaps with separate conditions devoted to different levels of intentional reversal of the figures might shed important light on what is indeed taking place in this interval. An important question here is whether a common mechanism up- or down-regulates switching in both visual and auditory tasks, and similarly, whether the locus of electrical activity related to such a regulation is frontal, in the system which controls, or more temporal or occipital, indicating that the control must first propagate to the sensory cortices themselves to modulate the sensation.

Second, since auditory bistability, tested with EEG is a new field, further investigation is certainly necessary to illuminate the scope and nature of the negative components observed here. It is recommended that many established paradigms from the literature on visual bistability be adapted to this new modality and used to highlight and further substantiate similarities and differences.

Conclusion

"A pure sensation [we see] to be an abstraction never realized in adult life. Any quality of a thing which affects our sense organs does also more than that: it arouses processes in the hemispheres which are due to the organization of that organ by past experiences, and the result of which in consciousness are commonly described as ideas which the sensation suggests. The first of these ideas is that of the thing to which the sensible quality belongs. The consciousness of particular material things present to sense is nowadays called perception. The consciousness of such things may be more or less complete; it may be of the mere name of the thing and its other essential attributes, or it may be of the thing's various remoter relations. It is impossible to draw any sharp line of distinction between the barer and the richer consciousness, because the moment we get beyond the first crude sensation all our consciousness is a matter of suggestion, and the various suggestions shade gradually into each other, being one and all products of the same psychological machinery of association. In the directer consciousness fewer, in the remoter more, associative processes are brought into play. [43]."

The goal of this thesis was to explore new stimuli and modalities with established procedures in cognitive neuroscience. Stimuli were designed, refined, and tested which were bistable but which differed from previous bistable stimuli in that they were dynamic and depended on a pairing of two consecutively presented frames in order to generate the illusion. Of these, one was a visual stimulus based on apparent motion research, and the other was a complex tone pair, utilizing the literature on the tritone paradox. These stimuli were tested with EEG, and compared to the prior literature regarding bistability.

Several components (an RN and an LPC for both visual and auditory modalities) were found which matched prior reports closely, although some of these components did not reach significance. New components were observed which are believed to relate to intention or to the anticipation of a perceptual reversal, and these components (central positivity in the visual condition and central negativity in the auditory) are highly significant and represent a promising avenue for future investigation.

The limits of sensation and perception are a field of psychology with far-reaching influences into philosophy and consciousness research. The question of what occurs in the brain at the moment when an identical stimulus is suddenly perceived differently is not resolved. Continued investigation requires continuous innovation from

studies such as the combined Rubin vase/Mcgurk effect paradigm of Munhall et. al. [66]. Another productive avenue is comparison between modalities to determine what characteristic responses are sensory and which are perceptual. The consequence of research in these areas is both a better understanding of the workings of these highly evolved systems by which human interface with the world, but also a small step toward understanding the processes of conscious experience and how it relates (or doesn't relate) to the world around us.

Appendix A

To Put Things in Perspective
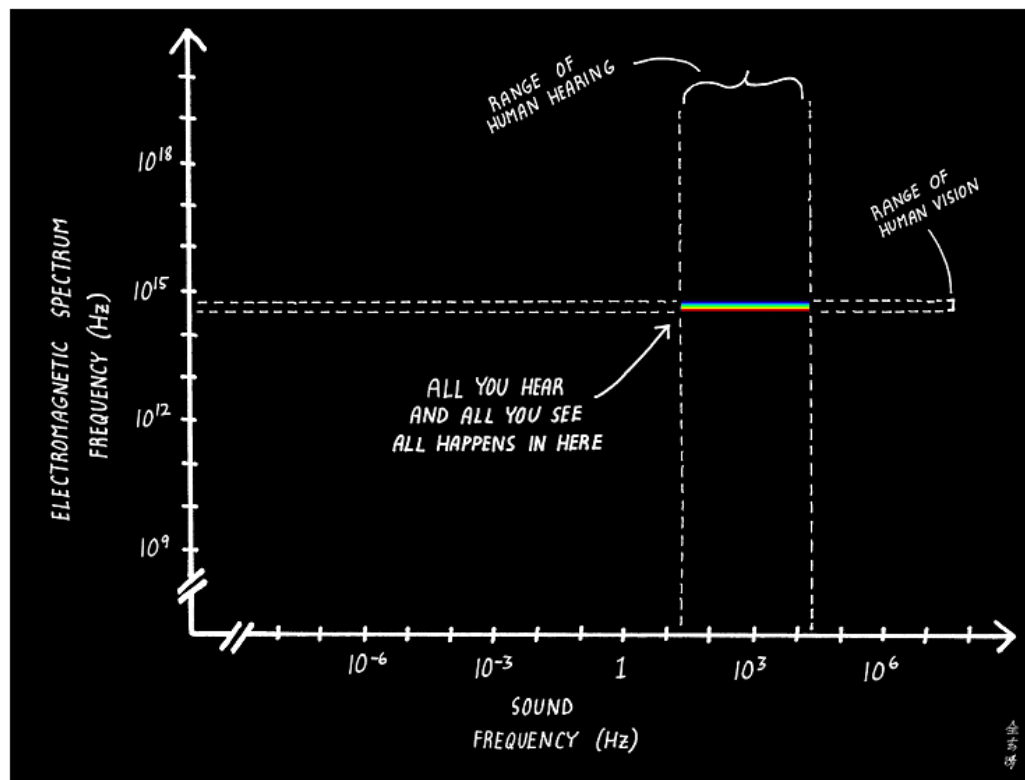
Figure A.1: If the doors of perception were cleansed every thing would appear to man as it is, infinite. For man has closed himself up, till he sees all things through narrow chinks of his cavern. –William Blake "The Marriage of Heaven and Hell" (Img. Source: Abstruse Goose).

References

[1] Sinauer Associates. Sensation & perception: Types of motion, 2009.

[2] F. Attneave. Multistability in perception. *Scientific American; Scientific American*, 1971.

[3] C. Baå-Eroglu, D. Strüber, M. Stadler, P. Kruse, and E. Baå. Multistable visual perception induces a slow positive eeg wave. *International Journal of Neuroscience*, 73(1-2):139–151, 1993.

[4] M.A. Bee and G.M. Klump. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *Journal of Neurophysiology*, 92(2):1088–1104, 2004.

[5] B. Bermond. Consciousness or the art of foul play. *Journal of Agricultural and Environmental Ethics*, 10(3):227–247, 1997.

[6] R. Blake. A neural theory of binocular rivalry. *Psychological review*, 96(1):145, 1989.

[7] F. H. Bool, J. R. Kist, J. L. Locher, and F. M. Wierda. *M. C. Escher: His Life and Complete Graphic Work.* Abrams, New York, NY, 1982.

[8] R.T. Born and D.C. Bradley. Structure and function of visual area mt. *Annu. Rev. Neurosci.*, 28:157–189, 2005.

[9] K. Bötzel and O.J. Grüsser. Electric brain potentials evoked by pictures of faces and non-faces: a search for face-specific eeg-potentials. *Experimental Brain Research*, 77(2):349–360, 1989.

[10] K. Bötzel, S. Schulze, and S.R.G. Stodieck. Scalp topography and analysis of intracranial sources of face-evoked potentials. *Experimental Brain Research*, 104(1):135–143, 1995.

[11] A.S. Bregman. Auditory scene analysis, 1990.

[12] A.S. Bregman. *Auditory scene analysis: The perceptual organization of sound.* The MIT Press, 1994.

[13] J. Britz, T. Landis, and C.M. Michel. Right parietal brain activity precedes perceptual alternation of bistable stimuli. *Cerebral Cortex*, 19(1):55–65, 2009.

[14] J. Britz, M.A. Pitts, and C.M. Michel. Right parietal brain activity precedes perceptual alternation during binocular rivalry. *Human brain mapping*, 32(9):1432–1442, 2011.

[15] The M.C. Escher Company B.V. M.c. escher: The official website, 2011.

[16] E.C. Cherry. Some experiments on the recognition of speech, with one and with two ears. *Journal of the acoustical society of America*, 25(5):975–979, 1953.

[17] R. Cusack. The intraparietal sulcus and perceptual organization. *Journal of cognitive neuroscience*, 17(4):641–651, 2005.

[18] D. Deutsch. A musical paradox. *Music Perception*, pages 275–280, 1986.

[19] D. Deutsch. The tritone paradox: Effects of spectral variables. *Attention, Perception, & Psychophysics*, 41(6):563–575, 1987.

[20] D. Deutsch. *Pitch class and perceived height- Some paradoxes and their implications.* Pendragon Pr, 1988.

[21] D. Deutsch. The tritone paradox: An influence of language on music perception. *Music Perception*, pages 335–347, 1991.

[22] D. Deutsch. Some new pitch paradoxes and their implications. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 336(1278):391–397, 1992.

[23] D. Deutsch. The tritone paradox: Some further geographical correlates. *Music Perception*, pages 125–136, 1994.

[24] D. Deutsch. Mothers and their children hear a musical illusion in strikingly similar ways. *The Journal of the Acoustical Society of America*, 99:2482, 1996.

[25] D. Deutsch. The tritone paradox: A link between music and speech. *Current Directions in Psychological Science*, 6(6):174–180, 1997.

[26] D. Deutsch. Mothers and their offspring perceive the tritone paradox in closely similar ways. *Archives of Acoustics*, 32(3):517, 2007.

[27] D. Deutsch. Tritone paradox, 2009.

[28] D. Deutsch et al. Paradoxes of musical pitch. *Scientific American*, 267(2):88–95, 1992.

[29] D. Deutsch, T. Henthorn, and M. Dolson. Speech patterns heard early in life influence later perception of the tritone paradox. *Music Perception*, 21(3):357–372, 2004.

[30] D. Deutsch, W.L. Kuyper, and Y. Fisher. The tritone paradox: Its presence and form of distribution in a general population. *Music Perception*, pages 79–92, 1987.

[31] D. Deutsch, T. North, and L. Ray. The tritone paradox: Correlate with the listener's vocal range for speech. *Music Perception*, pages 371–384, 1990.

[32] RJ Dolan, GR Fink, E. Rolls, M. Booth, A. Holmes, RSJ Frackowiak, and K.J. Friston. How the brain learns to see objects and faces in an impoverished context. *Nature*, 389(6651):596–598, 1997.

[33] J. Driver. A selective review of selective attention research from the past century. *British Journal of Psychology*, 92(1):53–78, 2001.

[34] F.H. Duffy, M.S. Albert, G. McAnulty, and A.J. Garvey. Age-related differences in brain electrical activity of healthy subjects. *Annals of neurology*, 16(4):430–438, 1984.

[35] P.A. Fuchs, C.J. Plack, and A. Rees. *The Oxford Handbook of Auditory Science: Hearing*, volume 3. Oxford Univ Pr, 2010.

[36] B. Fugiel. Waveform circularity from added sawtooth and square wave acoustical signals. *Music Perception*, 28(4):415–424, 2011.

[37] C.D. Geisler. *From sound to synapse: physiology of the mammalian ear*. Oxford University Press, USA, 1998.

[38] K. Grill-Spector, R. Sayres, and D. Ress. High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nature neuroscience*, 9(9):1177–1185, 2006.

[39] A. Gutschalk, C. Micheyl, J.R. Melcher, A. Rupp, M. Scherg, and A.J. Oxenham. Neuromagnetic correlates of streaming in human auditory cortex. *The Journal of Neuroscience*, 25(22):5382–5388, 2005.

[40] J.V. Haxby, M.I. Gobbini, M.L. Furey, A. Ishai, J.L. Schouten, and P. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001.

[41] P.J. Holcomb and H.J. Neville. Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, 5(4):281–312, 1990.

[42] T. Inui, S. Tanaka, T. Okada, S. Nishizawa, M. Katayama, and J. Konishi. Neural substrates for depth perception of the necker cube; a functional magnetic resonance imaging study in human subjects. *Neuroscience letters*, 282(3):145–148, 2000.

[43] W. James. The principles ofpsychology. *New York: Holt*, 2, 1890.

[44] R. Kanai, B. Bahrami, and G. Rees. Human parietal cortex structure predicts individual differences in perceptual rivalry. *Current biology*, 20(18):1626–1630, 2010.

[45] R. Kanai, D. Carmel, B. Bahrami, and G. Rees. Structural and functional fractionation of right superior parietal cortex in bistable perception. *Current Biology*, 21(3):R106–R107, 2011.

[46] N. Kanwisher, J. McDermott, and M.M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11):4302–4311, 1997.

[47] N. Kilian-Hütten, G. Valente, J. Vroomen, and E. Formisano. Auditory cortex encodes the perceptual interpretation of ambiguous sound. *The Journal of Neuroscience*, 31(5):1715–1720, 2011.

[48] D.B. King and M. Wertheimer. *Max Wertheimer and gestalt theory*. Transaction Pub, 2007.

[49] A. Kleinschmidt, C. Büchel, S. Zeki, and RSJ Frackowiak. Human brain activity during spontaneously reversing perception of ambiguous figures. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1413):2427–2433, 1998.

[50] C. Koch. *The quest for consciousness: A neurobiological approach*. Roberts & Company Publishers, 2004.

[51] W. Köhler and H. Wallach. Figural after-effects. an investigation of visual processes. *Proceedings of the American Philosophical Society*, 88(4):269–357, 1944.

[52] H.M. Kondo, N. Kitagawa, M.S. Kitamura, A. Koizumi, M. Nomura, and M. Kashino. Separability and commonality of auditory and visual bistable perception. *Cerebral Cortex*, 2011.

[53] J. Kornmeier and M. Bach. The necker cube–an ambiguous figure disambiguated in early visual processing. *Vision research*, 45(8):955–960, 2005.

[54] J. Kornmeier and M. Bach. Bistable perceptionalong the processing chain from ambiguous visual input to a stable percept. *International Journal of Psychophysiology*, 62(2):345–349, 2006.

[55] J. Kornmeier and M. Bach. Ambiguous figures–what happens in the brain when perception changes but not the stimulus. *Frontiers in Human Neuroscience*, 6, 2012.

[56] J. Kornmeier, W. Ehm, H. Bigalke, and M. Bach. Discontinuous presentation of ambiguous figures: How interstimulus-interval durations affect reversal dynamics and erps. *Psychophysiology*, 44(4):552–560, 2007.

[57] D.A. Leopold and N.K. Logothetis. Multistable phenomena: changing views in perception. *Trends in cognitive sciences*, 3(7):254–264, 1999.

[58] D.J. Levitin. *This is your brain on music: The science of a human obsession.* Dutton Adult, 2006.

[59] N.K. Logothetis. Single units and conscious vision. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1377):1801–1818, 1998.

[60] S.J. Luck. *An introduction to the event-related potential technique.* MIT Press, Cambridge, MA, 2005.

[61] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264:746 – 748, 1976.

[62] V. Menon, J.M. Ford, K.O. Lim, G.H. Glover, and A. Pfefferbaum. Combined event-related fmri and eeg evidence for temporal-parietal cortex activation during target detection. *Neuroreport*, 8(14):3029, 1997.

[63] M.P. Meyer. *Restoration of Motion Picture Film. Butterworth-Heinemann Series in Conservation and Museology.* Elsevier Science & Technology, 2000.

[64] C. Micheyl, B. Tian, R.P. Carlyon, and J.P. Rauschecker. Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, 48(1):139–148, 2005.

[65] K.A. Moores, C.R. Clark, J.L.M. Hadfield, G.C. Brown, D.J. Taylor, S.P. Fitzgibbon, A.C. Lewis, D.L. Weber, and R. Greenblatt. Investigating the generators of the scalp recorded visuo-verbal p300 using cortically constrained source localization. *Human brain mapping*, 18(1):53–77, 2003.

[66] KG Munhall, MW ten Hove, M. Brammer, and M. Paré. Audiovisual integration of speech in a bistable illusion. *Current Biology*, 19(9):735–739, 2009.

[67] I.R. Murray and J.L. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93:1097, 1993.

[68] L.A. Necker. 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. *The London and Edinburgh Philosophical Magazine and Journal of Science Series 3*, 1(5):329–337, 1832.

[69] BF O'Donnell, T. Hendler, and NK Squires. Visual evoked potentials to illusory reversals of the necker cube. *Psychophysiology*, 25(2):137–143, 1988.

[70] J. Orbach, D. Ehrlich, and A.H. Helen. Reversibility of the necker cube: I. an examination of the concept of" satiation of orientation". *Perceptual and motor skills*, 17(2):439–458, 1963.

[71] J. Orbach, E. Zucker, and R. Olson. Reversibility of the necker cube: Vii. reversal rate as a function of figure-on and figure-off durations. *Perceptual and Motor Skills*, 22(2):615–618, 1966.

[72] L.S. Penrose and R. Penrose. Impossible objects: A special type of visual illusion. *British Journal of Psychology*, 49(1):31–33, 1958.

[73] J.R. Pierce. *The nature of musical sound.* Academic Pr, 1999.

[74] Walter Piston. *Harmony.* W.W. Norton Co., New York, NY, 1941.

[75] M.A. Pitts, A. Martínez, and S.A. Hillyard. When and where is binocular rivalry resolved in the visual cortex? *Journal of Vision*, 10(14), 2010.

[76] M.A. Pitts, A. Martínez, C. Stalmaster, J.L. Nerger, and S.A. Hillyard. Neural generators of erps linked with necker cube reversals. *Psychophysiology*, 46(4):694–702, 2009.

[77] M.A. Pitts, J.L. Nerger, and T.J.R. Davis. Electrophysiological correlates of perceptual reversals for three different types of multistable images. *Journal of Vision*, 7(1), 2007.

[78] R. Plomp. Pitch of complex tones. *The Journal of the Acoustical Society of America*, 41:1526, 1967.

[79] D. Pressnitzer and J.M. Hupé. Is auditory streaming a bistable percept. In *Forum Acusticum, Budapest*, pages 1557–1561, 2005.

[80] D. Pressnitzer and J.M. Hupé. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, 16(13):1351–1357, 2006.

[81] J. Qiu, D. Wei, H. Li, C. Yu, T. Wang, and Q. Zhang. The vase-face illusion seen by the brain: An event-related brain potentials study. *International Journal of Psychophysiology*, 74(1):69–73, 2009.

[82] F. Ragozzine and D. Deutsch. A regional difference in perception of the tritone paradox within the united states. *Music Perception*, pages 213–225, 1994.

[83] D. Regan. *Human perception of objects: early visual processing of spatial form defined by luminance, color, texture, motion, and binocular disparity.* Sinauer Associates, 1999.

[84] J. C. Risset. Paradoxes de hauteur: Le concept de hauteur sonore n'est pas le meme pour tout le monde. *Seventh International Congress of Acoustics, Budapest*, 10:20, 1971.

[85] E. Rubin. Figure and ground. *Readings in perception*, pages 194–203, 1958.

[86] M. Sato, M. Baciu, H. Loevenbruck, J.L. Schwartz, M.A. Cathiard, C. Sege-barth, and C. Abry. Multistable representation of speech forms: A functional mri study of verbal transformations. *NeuroImage*, 23(3):1143–1151, 2004.

[87] B.M.A. Sayers, HA Beagley, and WR Henshall. The mechanism of auditory evoked eeg responses. *Nature*, 1974.

[88] JF Schouten. *5 [five] Articles on the Perception of Sound (1938-1940).* Instituut voor Perceptic Caderzoek, 1960.

[89] R.N. Shepard. Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America*, 36:2346, 1964.

[90] R.N. Shepard. Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89(4):305, 1982.

[91] P. Sterzer, A. Kleinschmidt, and G. Rees. The neural bases of multistable perception. *Trends in cognitive sciences*, 13(7):310–318, 2009.

[92] D. Strüber and C.S. Herrmann. Meg alpha activity decrease reflects destabi-lization of multistable percepts. *Cognitive Brain Research*, 14(3):370–382, 2002.

[93] M.J. Tarr, I. Gauthier, et al. Ffa: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature neuroscience*, 3:764–770, 2000.

[94] M.J. Tovée. *An introduction to the visual system.* Cambridge Univ Pr, New York, 2008.

[95] A.M. Treisman. Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12(4):242–248, 1960.

[96] P. v. Schiller. Stroboskopische alternativversuche. *Psychological Research*, 17(1):179–214, 1933.

[97] D.C. Van Essen, C.H. Anderson, and D.J. Felleman. Information processing in the primate visual system: an integrated systems perspective. *Science*, 255(5043):419, 1992.

[98] L.P.A.S. van Noorden. *Temporal coherence in the perception of tone sequences.* Institute for Perceptual Research, 1975.

[99] Y.A. Vasilenko and LB Shestopalova. Moving sound source discrimination in hu-mans: Mismatch negativity and psychophysics. *Human Physiology*, 36(2):139–146, 2010.

[100] N.J. Wade. Early studies of eye dominances. *Laterality: Asymmetries of Body, Brain and Cognition*, 3(2):97–108, 1998.

[101] W.D. Ward. *Absolute Pitch.* Academic Pr, 1998.

[102] R.M. Warren and R.L. Gregory. An auditory analogue of the visual reversible figure. *The American Journal of Psychology; The American Journal of Psychology*, 1958.

[103] L. Weiskrantz. *Blindsight: A case study and implications*, volume 12. Oxford University Press, USA, 1990.

[104] N. Zaretskaya, A. Thielscher, N.K. Logothetis, and A. Bartels. Disrupting parietal function prolongs dominance durations in binocular rivalry. *Current Biology*, 20(23):2106–2111, 2010.