Transformation, Restoration and Migration. Illusions in the Perception of Speech, Music and other Complex Sounds.

A dissertation submitted in fulfilment of the requirements

of the degree of Doctor of Philosophy.

The City University, London.

by Peter Mayer

April 1994

Acknowledgements

I wish to acknowledge my debt and offer my thanks to the following people without whose help and encouragement this thesis would never have been submitted. On a practical level I would like to thank I. K. Ong and all those at Skoob Books whose frienship and support were vital in the completion of this research. I am also indebted to Dr Ian Cross who generously gave his advice and much needed assistance in the construction and selection of musical stimuli. However, academically I owe the most to Dr Zofia Kaminska whose patient and insightful supervision of this thesis has been utterly invaluable. At a personal level the unselfish, unstinting support and understanding of my wife Chris Mayer has been the single most important factor in ensuring that words were turned into deeds.

ABSTRACT

This thesis questions the traditional view of speech perception as a unique form of auditory processing by exploring certain parallels between the cognitive processing of speech and other auditory stimuli.

Three phenomena singularly associated with speech perception - verbal transformation, phonemic restoration and click migration - form a tri-partite basis of investigation. Each phenomenon is examined in turn in a series of experimental investigations and in each case evidence for a non-verbal analogue is found.

Part 1 demonstrates that the verbal transformation effect obtains in both music and other complex non-verbal sounds, belying its original name.

Part 2 finds that illusory continuity in music and other complex non-verbal sounds mirrors the function of phonemic restoration in speech.

Part 3 highlights cognitive structuring in music, as revealed by click migration phenomenon, showing it to be closely analogous to that found in verbal material.

The theoretical basis of each phenomenon is reassessed in the light of the experimental findings and the implications for auditory processing are examined.

CONTENTS

General Introduction

The auditory world has traditionally been divided by cognitive psychologists into two categories : speech and other sounds. Speech perception has been characterised as a unique form of auditory processing, a singular domain in which the final percept bears scant relation to the physical waveform. In contrast, the perception of sounds which constitute the rest of the auditory domain was deemed to be informed by a direct and open relationship with the external world. This dichotomous world has been drawn with good reason. The evidence for the uniqueness of speech sounds is formidable, converging from several directions.

Neurology and neuropsychology point to the usually unilateral, typically left-hemispheric and precise specialisation for, and localisation of, speech processing (McCarthy and Warrington, 1990). Speech sounds are also unique in that their perception is intertwined with their production. Evidence from electrical stimulation mapping of the brain at the level of individual neurones implicates the same neurones in both discrimination and articulation of phonemes (Ojemann, 1982), while visual information defining another sound to produce an auditory percept representing neither (McGurk and MacDonald, 1976), again implicating cross-referencing between input and output pathways.

A further, yet indirect, source of influence on the 'speech is special' argument has spread from the domain of psycholinguistics. This field of study has historically emphasised the unique nature of human language as a form of communication. Hence it was assumed that in order to decode and reintegrate the complexities of linguistic structure from the auditory stream it was necessary to posit the existence of dedicated and highly specialised speech processing systems. For example, Liberman (1982) states, when discussing speech sound processing mechanisms, "The distinctively linguistic function of these specializations is to provide for efficient perception of phonetic structures that can also be efficiently produced." Another aspect of linguistic influence can be discerned in the analysis of segmentation processes, which have concentrated on the psychological instantiation of grammatical units extracted from the speech stream (Flores D'Arcais and Schreuder, 1983).

However, the most pervasive and persuasive characteristic of speech distinguishing it from other sounds is the lack of an invariant relationship between the acoustic signal which serves as the auditory stimulus in perception, and the way in which that stimulus is actually perceived.

This lack of correspondence between acoustic characteristics and emergent percept is observable at different levels of analysis. At the acoustic level, a speech sound basically consists of two (sometimes three) fundamental bands of energy at, or sweeping across, particular frequencies. In the wider auditory world such bands are heard as two sounds at a steady, or changing, pitch in a fairly lawful manner.

Consonance between the physical and psychological world appears to be preserved. But speech percepts do not manifest such a relationship with external progenitors. The psychological reality of speech sounds seems to stem from a more obscure mode of translation from the physical, and the bands of energy are heard as a single sound, a unified entity - a vowel or a consonant. (Liberman et al. 1967, Foss and Bank, 1982, Handel, 1989).

Hence, the processes that control the perception of speech have been characterised as uniquely creative and constructive. In support of this contention a number of illusory phenomena have been observed surrounding speech perception. The theoretical interpretations of all of these phenomena are based on the same premise, that speech processing is qualitatively different from all other auditory domains. Therefore if it is possible to establish a range of non-verbal analogues for these essentially verbally bound illusions, then the unique nature of the processes which govern speech fall into question.

It is the intention in this thesis to examine three illusory phenomena which have been used to illustrate the unique nature of the processes which decode and reintegrate speech : verbal transformation, restoration, and click migration. Verbal transformations may be described as the illusory changes in percept which occur when a stimulus word or phrase is presented repeatedly. Restoration is the process whereby the perception of continuity is preserved when a small segment of a stimulus is removed and replaced with noise, while the click migration phenomenon is the tendency of an extraneous noise to be perceptually mislocated at a position in the stimulus which marks a segmental boundary.

It is the aim of the experimental studies and theoretical discussions which follow to examine non-verbal analogues for these phenomena and explore the implications for some of the processes involved in perception across different auditory domains. In order to achieve the objective of this thesis it is divided into three parts, each dealing with a different illusory phenomenon. The purpose of each part will be to consider the literature and theory which is associated with a separate illusion, presenting experimental evidence to explore the case for a non-verbal analogue and considering the effects on how auditory processing systems are conceived. If non-verbal analogues can be demonstrated for each of the illusory phenomena then the case for asserting that speech processing is unique can be legitimately reassessed.

Therefore the primary aim of this thesis is to establish a set of non-verbal analogues for the three illusory phenomena under consideration: verbal transtormation, phonemic restoration and click migration. The secondary aim is to then examine the processing implications which would follow from the verification of these illusions beyond the verbal domain.

,,,,,

PART 1

The Not Just Verbal Transformation Effect

This Part of the thesis focuses on the phenomenon of illusory perceptual changes that occur when a single physical token of brief auditory stimulus is repeated continuously for a few minutes. Although the stimulus never physically alters in any respect, listeners routinely report gross distortions in their percept. Historically, words were the only stimuli used and the phenomenon was seen as a speech based effect. Explanatory accounts which centred on the uniqueness of speech processing mechanisms provided the most coherent and persuasive models.

The experiments presented and explanations suggested in this thesis seek to place the phenomenon in the wider auditory world and to analyse it within a more modem theoretical context. In doing so, they challenge established views.

CHAPTER 1 The Transformation Illusion

Early Research

Unlikely as it may seem in the light of subsequent work, the first experimental use of repeated verbal input can probably be attributed to the then pre-Behaviourist B F Skinner (1936). Skinner referred to the repetition of vague and deliberately indistinct speech sounds as "a sort of verbal ink blot". Indeed, he used this form of stabilised auditory input as a kind of apperceptive test called the "verbal summator". The procedure entailed subjects listening to repeated presentations of a speech stimulus which was constructed to be both phonologically and auditorily indistinct. Subjects invariably claimed to be able to make sense of these sounds. They often reported that the recorded voice had made comments concerning their private lives. As soon as the subject had made any response, the recording was switched off and another indistinct set of speech sounds was played.

Skinner viewed the phenomenon not as an auditory illusion but as the perceptual imposition of organised structure on disorganised information, much in the same way as one might pick out familiar shapes in the patterns of an ink blot drawing. However, as Warren (1983) has pointed out, the ink blot viewer does not imagine new sections of ink whereas the listeners to repeated speech sounds imagined new words emerging from the indistinct auditory fragments. Even though this "verbal summator" did not flourish as a projective test it is important to note, in the light of arguments that are to be made here at a later stage, that the theory which explained its function clearly assumed the subject to be making "associations" between disorganised ambiguous input and existing mental representations. In short, the "summations" were theorised as the product of a complex associative process.

The Necker Cube Analogy

The use of stabilised auditory input as a form of projective testing did not continue for long but the resurgence of interest in cognitive psychology in the late 1950's brought the next contribution to the area, when Warren and Gregory (1958) claimed to have found a verbal analogue of the visual reversible figure.

It had been known for a long time that, when disambiguating cues are removed, visual figures can be created which are open to more than one meaningful interpretation. The most famous example of this is the Necker cube. Warren and Gregory reasoned that words were also disambiguated by their context and, once isolated and presented in continuous repetition, they might also be reversible, much in the same way as the Necker cube alternates from one plausible interpretation to another. The example they used was the word 'SAY' which they reported would "shift abruptly to 'ACE' and back again" thus performing a reversal in an auditory mode analogous to the visual reversible figure. Warren and Gregory used the term "verbal alternation" to describe the effect and noted that it seemed similar enough in principle to visual reversible figures to suggest that the same basic perceptual rules underpinned both illusions.

Tempting though this analogy was, the notion of an auditory equivalent of the Necker cube did not prevail for long. Four areas of divergence became clear:-

(i) The visual illusion consists of an alternation between two forms whereas the auditory illusion generally induces many different forms.

(ii) The transitions of the visual illusion do not involve significant distortions of the figure and the alternatives are predictable whereas the auditory illusions frequently produce gross distortions and completely new words.

(iii) The visual illusion provokes a limited number of identical perceptual interpretations common to all subjects whereas the illusory forms perceived in the auditory mode vary greatly from subject to subject.

(iv) The visual illusion is restricted to special configurations, whilst the auditory illusion is not restricted to any particular word.

Warren (1961a) was the first to discuss these differences. As experimentation on stabilised auditory inputting increased, so visual analogies as explanatory models decreased in importance.

The Verbal Transformation Effect

The term "Verbal Transformation (VT) effect" was first coined by Warren (1961). This study constituted the first systematic attempt to unravel the important variables which influence the perception of illusory changes in repeated presentations of a verbal stimulus. In a series of experiments, Warren varied both qualitative and quantitative aspects of the stimulus. Using a variety of speech sounds, ranging from a single word to a short sentence, he manipulated both the rate at which the stimulus was presented and the quality of the listening conditions from faint and masked to loud and clear.

The results obtained by Warren in his study can be summarised as follows:-

(i) Simple stimuli (e.g. a one syllable word) tend to elicit most illusory changes whilst complex stimuli (e.g. a whole phrase) remain closer to their original form on repeated presentation.

(ii) The number of stimulus repetitions, rather than exposure duration, is the salient factor in producing new forms. That is, a stimulus will induce more forms if it is played in continuous repetition (e.g. with no inter-stimulus interval) than if it is presented at a slower rate (e.g. with a small inter-stimulus interval) for the same period of time.

(iii) An indistinct or masked stimulus leads to a lower rate of VT production than a clear and distinct one.

(iv) Once the stimulus is loud enough to be heard clearly, increases in intensity do not affect the rate of VT production.

This work was effectively extended by Warren and Warren (1966) using the same stimuli to test a wide age range of subjects. VT's were found to vary as a function of age. At age 5 years and below children did not report VT's; at age 6 years approximately 50% of those tested reported hearing some illusory changes; by 8 years all the children heard some VT's. The rate of VT's observed at this stage remained constant through the rest of the age ranges tested and only showed any marked decline in subjects of 60 years and over.

The theories and explanations that have been proposed to account for the VT effect will be examined in the following section. For now it is enough to note that the two studies mentioned above may be seen as the dividing line between the early research and the more serious consideration of this form of auditory illusion as an area of study in its own right.

Theories of the Verbal Transformation Effect

To date, there is no satisfactory theoretical account of all the illusory phenomena associated with VT's. Theories so far advanced may be divided into three categories:-

1. Those which view VT's as a phenomenon analogous to the stabilised retinal image.

2. The category of models which seek to explain VT's in terms of fatigue.

3. The theory of VT's as a reflection of constructive speech processing mechanisms.

These approaches will be examined in turn :-

1.Stabilised retinal image analogy.

The view that VT's are an auditory effect which corresponds to the stabilised retinal image essentially grew out of the failure to establish a direct analogue between it and the Necker cube. In an attempt to pursue a visual analogy, an alternative visual effect was proposed as the basic model.

It has long been known that if one stares at a word "the meaning evaporates and the symbols seem unfamiliar" Severance and Washburn (1907). This procedure does not produce any transformations because of physiological nystagmus - continuous involuntary eye tremor which has the effect of producing a constantly mobile image, so that at the retinal level no cell receives unchanging stimulation. However, it is possible to cancel out the effects of eye movements with appropriate optical systems (Ditchburn and Fender 1955) and create a stabilised retinal image. Further experiments using such techniques found that when an unchanging pattern of visual stimulation is maintained, portions of the stabilised visual image fade and reappear with a tendency for meaningful perceptual organisations to dominate. This produces changes which certainly are in some respects similar to VT's. For example, when the word "BEER" was used as a stimulus, subjects reported a series of changes involving whole letters or parts of letters i.e. PEER, PEEP, BEE and BE. These resemble VT's to the extent that the changes tend to "preserve or create meaning" (Warren 1968).

The idea that changes which occur under conditions of a stabilised retinal image and VT's are both reflections of the same mechanism was to some extent supported by Swales and Evans (1967) who showed certain similarities between the two effects. The most important common feature was that both VT's and stabilised retinal image changes occurred more quickly for complex than for simple stimuli.

Although this evidence offered some prima facie support for a link between the two phenomena, the 'analogy' approach failed to gain any widespread recognition as differences still outweighed similarities. The most basic difference is that the 'continuous' presentation of a stabilised retinal image cannot be entirely analogous to the 'repeated' presentation of the auditory stimulus. In this sense, the only true auditory analogue would be the continuous presentation of a tone or tones. Also, changes reported in stabilised retinal images, while preserving meaning, tend to be of a subtractive nature whereas VT's very often involve a reorganisation of the stimulus word or the production of a completely different word and hence introduce an additive influence.

For example, Warren (1961a) used the word "RIGHT" as a stimulus with one subject reporting as follows:-

"ripe, right, white, white-light, right, right-light, ripe, right, ripe, bright-light, right, ripe, bright-light, right, bright-light."

The range and creativity involved in the auditory illusion tends to underline the divergence on a qualitative level between the two phenomena. In short, the comparison between the two effects has not really produced any insight into the function of either illusion.

2.Fatigue models

These models focus on the possibility that VT's are caused by the operation of fatigue at one of two levels - (i) Auditory or (ii) Cognitive.

(i) Auditory Fatigue Model

This explanation is based on the idea that continued repetition of the same physical waveform may cause fatigue in a particular set of auditory receptors causing a distorted, or not quite accurate, signal to be fed into higher level processing. This theory has a great deal of intuitive appeal, relying as it does on erroneous auditory processing almost randomly accessing higher level representations and thus accounting for the range of observed illusory transformations. The plausibility of fatigue operating in this way to produce illusory percepts would probably make this explanation the preferred option of the intelligent layman.

The problem with this "common sense" theory is that it does not tally with critical aspects of the available evidence. First, if fatigue lay at the basis of the VT effect, the longer the stimulus continues to be repeated, the rate at which new forms are observed should increase. That is, the greater the fatigue, then the rate of perceptual production of new forms should be a positive function of duration of listening. Therefore, the greater the fatigue, the greater the probability of erroneous processing accessing different mental representations and more new forms should be produced by the listener. According to Warren (1961a) "the rate at which VT's were called out was the same for each minute of the three minute test period, but the rate of appearance of new forms decreased during each minute." This finding directly contradicts what might be expected if fatigue were the cause of the illusory percepts. In support of this position, Warren (1968) claimed that with a clear stimulus, that is, one that is not degraded or masked, basic audio-physical variables do not affect the rate or variety of changes whereas if cochlear or afferent neural fatigue were the primary agents involved then alterations in the physical nature of the stimulus would have some effect on the percept. So although auditory fatigue offers a possible explanation of how the range of illusory changes might have been produced, it fails as an hypothesis in the light of existing empirical evidence.

(ii) Cognitive Fatigue Theory

This theory was advanced by Calef et al (1974). At the centre of the model lies the idea that processors involved in speech perception do not maintain a sufficiently high level of arousal when presented repeatedly with the same information. Hence, when arousal level drops as a result of lack of variety in stimulation, errors occur in the system, representations of words other than the stimulus word are accessed and thus the percept is altered.

The difference between this model and the previous one is that the action of fatigue impinges at a higher, more "cognitive", level of the process. So, in this formulation, the semantic content of the stimulus would be a more salient factor in determining the arousal level than its physical characteristic. The evidence which Calef et al brought forward to support this hypothesis was the finding that taboo words (those which cause emotional arousal e.g. rape, sex etc.) produce significantly fewer VT's than did neutral words (e.g. rhyme, sketch etc.). Their explanation of this was that emotional words "produced and maintained a higher level of arousal" in the perceptual system than did neutral words. Hence, the greater number of illusory percepts experienced by subjects could be attributed to the low level of arousal produced by the neutral words.

Although this model offers a rudimentary explanation of how the VT effect might function, Calef et al failed to expand on how this system of arousal might work in detail. For example, would they propose some hierarchy of arousal for words of varying emotional context? In fact, the model has neither been extended by the authors of the original study nor picked up in any of the literature in the field. This may be attributable to the methodology used by Calef et al, which could have cast doubt on the validity of the findings. Stimuli were presented to subjects in a group study context and responses were written down. The group atmosphere would presumably encourage subjects (especially the more timid) to suppress their responses to words such as "rape" and "sex". This also may have been compounded by subjects who felt uncomfortable about responding at all to repetitions of such highly charged stimuli.

Perhaps due to methodological doubts and a lack of theoretical specificity, this model has not received any further attention or investigation but its relative anonymity can also be attributed to the success and wide degree of acceptance which has been achieved by the model to be outlined in the following section.

3. VT's as a reflection of constructive speech processing mechanisms.

This model sees the phenomenon of VT's as part of an overall approach the central tenet of which is that processes involved in speech perception are unique. The emphasis here changes from viewing VT's as an isolated effect requiring a specific interpretation to an explanation which explicitly links these illusory changes to a cluster of other auditory phenomena in speech. These phenomena include categorical perception, click migration and, more particularly, phonemic restoration. The common feature is the clear divergence between the physical reality of the auditory input and the reconstructed psychological reality of the percept.

The most mature articulation of this view may be found in Warren (1982) For Warren, the primary indication that VT's are a reflection of constructive speech processing mechanisms comes from the lack of response in the very young and the old. Warren asserts that "the reorganisation of repeated words is related to the reorganisation of connected discourse employed when a preliminary organisation of speech sounds into words and phrases is not confined by subsequent context." Hence, the reason why VT's are not observed in the very young is that these "reorganisational processes" do not appear until language skills have reached a certain stage of development. This stage, Warren hypothesises, occurs at age 6-7 years (Warren and Warren 1966). Furthermore, the decline in reported VT's in older subjects (60 years and over) is attributed to "adaptive changes in processing strategy" due to an increasing inability to perform tasks "involving the use of complex storage".

For Warren, the variation of VT's with age offers a background to a more specific description of the function of the VT effect in higher order reorganisational processes. He proposes that the VT effect is closely related to "mechanisms employed normally for the resolution of ambiguities and correction of errors" in everyday speech perception. In a confusing auditory environment, where sections of speech may be imperfectly received (e.g. a noisy room with loud machinery or with multiple conversations taking place) constructive and reorganisational processes must be engaged in creating the final percept from incomplete information. Warren suggests that VT's are a by-product of this system when it is confronted with repeated presentations of stimulus words that are out of the context of connected discourse. Warren further asserts that other auditory effects such as phonemic restoration, click migration and categorical perception are also facets of the same system.

Although these phenomena display different functions within the system, Warren thought it plausible to examine a conjunction between one of them, phonemic restoration, and the VT effect. Phonemic restoration describes a phenomenon whereby a noise is substituted for a segment of a word on a tape recording and subjects who hear the altered presentation report hearing a complete utterance plus an extraneous sound. The missing phoneme is effectively perceptually restored by the listener.

It was thought possible that an experimental link could be made between these two effects because both phenomena were the product of functionally creative processes. In phonemic restoration the missing segment is reinstated in the final percept and in VT's completely new words are heard. Both effects feature the active creation of aspects of the final conscious realisation which were not present in the original auditory signal.

Obusek and Warren (1973) pursued this line of enquiry by combining procedures producing the VT effect and phonemic restoration in one experiment. They took a stimulus word "MAGISTRATE", removed the "S" and replaced it with a noise burst. Using the new stimulus in a standard VT paradigm, they found that a significantly higher proportion of the observed illusory transformations occurred around the position of the missing phoneme than appeared in a control condition where the intact word was used as a stimulus in the VT paradigm.

Subjects reported significantly more VT's associated with the "S" in the missing condition even though they claimed to be perceiving the complete word. The experiment had succeeded in manipulating the production of illusory percepts by combining the two paradigms and theoretically linking the two phenomena. Warren concluded that phonemic restorations and VT's "operate together to minimise errors and enhance the intelligibility while listening to speech under noisy conditions."

This explanation of the VT effect by Warren sets the phenomenon in a functional context by explicit linkage to other speech perception phenomena. This also, by implication, sets VT's among that group of auditory effects which are used as evidence for the more general proposition that speech processing mechanisms are unique.

The Associative Theory - A New Explanation of VT Effect

Warren's explanation of the VT effect gives it a place among constructive speech processing mechanisms, explicitly linking it to the creative aspects of phonemic restoration and, as such, advances it as a strand of evidence supporting the unique nature of speech perception. It is the intention here to introduce a fresh view of the VT effect which de-couples this auditory illusion from speech processing mechanisms in particular and from an exclusively verbal interpretation of the phenomenon in general.

The general model proposed here by the author contends that VT's are not the necessary product of any particular set of processing mechanisms but are the result of compelling associative stimulation caused by the repetitive accessing of the same mental representation. It may be that when a word is used as auditory input, then not only the stimulus itself is accessed but the whole domain of associated representations "at all levels" are stimulated. Since the early demonstrations of priming (Meyer and Schvaneveldt et al 1971) it has long been known that the recognition of a stimulus can be significantly facilitated by prior presentation of related input. This makes the point that any stimulus initiates a level of activation in a domain of related representations. Without drawing any further theoretical parallels with early priming studies, VT's can be explained as representations associated with the input word which have, through numerous repetitions, received sufficient stimulation to achieve conscious realisation. All levels of the processing system are activated under the precepts of this model, phonemic through to semantic. Hence, the activation of related phonemes may produce non-words or words only related phonologically while, on another level, semantic relationships between words may be entirely subject specific, thus giving an observer the appearance of a totally unrelated new form. A simple stimulus may therefore be able to activate an extremely large domain of associated representations. VT's are thus the reverberations, both vertical and horizontal, through stimulus related domains in an interconnected and distributed system.

The further theoretical implications of this approach will be explored with the presentation of new evidence later in this chapter. At this point, a re-examination of the existing evidence will show that the Associative model, as it can be called, offers a "better fit" in explanatory terms than the speech process model put forward by Warren.

The centrepiece of Warren's theory is the explicit link between VT's and other verbal phenomena. The basis for this line of argument stems from the study by Obusek and Warren (1973) which made a clear experimental connection between phonemic restorations and VT's by combining the two phenomena in one paradigm. The results clearly indicated that the illusory changes perceived by the subjects tended to cluster in a significantly disproportionate way around the missing phoneme of the stimulus word. This indicated to Warren that the same creative and reorganisational processes were involved in the functional realisation of both phenomena. But if the two phenomena really arise from the same speech processing mechanism, then it should be possible to find parallels between the conditions under which they operate and their effect on the final percept.

If, however, this examination is undertaken, differences rather than similarities emerge. For example, the conditions under which missing phonemes are restored preserve a highly restrictive context. The critical segments are removed from otherwise complete words which are usually part of a longer phrase or sentence. Thus, in this phenomenon, the context strictly regulates the final percept.

An experiment by Sherman (1971) demonstrates this clearly. It was found that a disambiguating word at the end of a sentence significantly altered the nature of the subject's perception of the missing phoneme. For example, in the phrase "The #eel came off the old ...." (where # denotes the phoneme replaced by a noise burst), the subject's perception of the missing speech segment was manipulated by ending the sentence with a number of different nouns each conveying a different meaning for the sentence e.g. "car", "shoe", or "orange". So when the stimulus ended with the words "....the old shoe" subjects reported hearing "the heel came off....". The perception of the missing phoneme was constrained by the subsequent context so that, when "car" was the disambiguator, "wheel" was heard as the critical word and "peel" was perceived when "orange" finished the sentence.

This evidence indicates that this phenomenon operates as a function of the context conditions present in the utterance. That is, the predictive and restrictive information contained in the stimulus forced the system to reconstruct the brief missing segment in order to make the distorted word consonant with its context, thereby preserving meaning in the final percept. The restorative process here operates under conditions that are highly constrained in order to maintain phonemic and semantic continuity.

In contrast, the processes which underlie the VT effect do not seem to relate to context at all with the effect that meaning, far from being preserved in the final percept, is often lost even to the extent that non-words are reported. Illusory changes perceived by subjects do not maintain phonemic and semantic continuity and, by their very nature, they are at variance with the original stimulus. In linking the two phenomena, Warren is theorising that essentially restorative and reconstructive processes which serve to preserve meaning also so "reorganise" other stimuli that all sense is lost. The conditions under which these phenomena operate and their effect on the final percept seem to be so different that the only basis for a link between them lies in the experimental findings of Obusek and Warren (1973). The validity of this result is not in question, but a different interpretation can be placed on the findings, one which offers a more plausible rendering of the known evidence.

It is necessary to view phonemic restoration (PR) and VT as essentially different sorts of phenomena. PR should be seen as an adaptive reconstructive "mechanism" operating normally under experimental conditions while VT's should not be seen as a "mechanism" at all, but rather as the product of subversion of the normal function of the perceptual system with a stimulus configuration that could only be encountered in the laboratory. This is not intended to take the form of a value judgement as to whether one phenomenon or the other is more worthy of study. Both offer insights into the way that our final conscious percepts are created, but the status of each phenomenon is vital if to a clear understanding.

In bare terms, PR is an illusion generated by the perceptual system in order to preserve the continuity of the percept whilst VT's are stimulus-generated illusions caused by putting the same system under unique pressure. The Obusek and Warren results can therefore be explained in different terms.

The operation of PR relies on the context of the missing segment in order to reinstate it accurately. However, before the missing segment has been completely disambiguated, the perceptual system must be prepared for the incomplete token to end in a number of possible ways. Thus it would be necessary for the perceptual system to prime a set of possible options as the stimulus unfolds in time, each one close to the level of conscious realisation, in order to create the instantaneous sensation experienced by subjects as the word is disambiguated by the final segment. This high level of systematic priming serves to enhance the associative priming generated by repeated presentations of the same stimulus. The adaptive activation caused by the system's restorative mechanisms would not, under normal conditions, have any perceptual repercussions. But, when the token is represented over and over again, the increased stimulation to the related representations raises the probability that a connected item will be primed often enough to induce a new conscious percept.

Whereas Warren's explanation of the original finding concluded that the increase in VT's associated with the missing segment was due to the action of two speech specific processing mechanisms working together, the reassessment of the findings presented here in what can be termed the 'Associative model' sees the two phenomena as essentially different. More precisely, the associative model views the two as coincident but springing from different causes. As the stimulus is repeated, PR keeps on performing its adaptive function, thereby maintaining a consonant and continuous percept, but its action under these repetitive conditions only serves to channel stimulation to associated interconnected representations. This effect makes the occurrence of a transformation more likely as it restricts the number of primed items but increases the stimulation they receive. To understand this notion it is necessary to remember that VT's are characterised here, not as part of the normal speech processing mechanisms, but as creations of a specific experimental situation. In short, VT's do not appear in real life, they are the product of laboratory manipulations and a phenomenon whose action is facilitated by the adaptive functioning of PR - a phenomenon which does have a place in normal processing.

The above arguments suggest therefore that the findings of Obusek and Warren (1973) reflect a real effect but that their explanation is flawed. Evidence to support this contention, and the new hypothesis that transformations are an associative phenomenon to be found in other auditory domains, will be described during the further course of this part of the thesis.

A Non-Verbal Analogue of the VT Effect

The evidence for a non-verbal analogue for the VT effect is extremely sparse. Although early experiments by Elliot (1963), Taylor and Hemming (1963), Fenelon and Blayden (1968) and Perl (1970) had investigated the effect of non-verbal stimuli in repeated presentations, Obusek and Warren (1973) were still able to say with confidence that "there is as yet no convincing evidence that perceptual transformations analogous to VT's exist for repeated sequences of non-speech sounds." This, on balance, was a justified assessment as the studies quoted offered an extremely narrow basis of comparison. Elliot employed repeated noise bursts, Taylor and Hemming used a pattern of beeps whilst Perl, Fenelon and Blayden used pure tones. None of the experiments included verbal stimuli as a comparative condition within the same procedural manipulations. Thus, since some studies encountered only certain types of behaviour whilst others recorded number of changes without reference to type, coherent analysis in order to establish a non-verbal analogue was impossible.

The most convincing evidence for the existence of illusory changes in non-verbal stimuli came a little later on from Lass, West and Taft (1973). Their experiment offered a broader base of stimulus comparison (although still no verbal conditions); three pure tone categories (250, 1000 and 4000 Hz), a broad band white noise and a musical motif were used. All five experimental conditions produced some transformations, with the musical motif being clearly the most potent. This latter type of stimulus required by far the fewest repetitions before the first transformation was reported, with the number of new forms being roughly equivalent to the average number of new forms produced over a range of repeated words in Warren's experiment (1961a). Lass et al also favourably compared the performance of the musical motif to speech stimuli which were used in a prior experiment conducted under similar listening conditions (Lass and Golden 1971). The response classification procedure was also the same in these two studies, thus giving more credence to the comparison between the musical condition and the speech sounds.

It should be emphasised however that this finding cannot be considered to constitute adequate evidence of a non-verbal analogue as only one musical motif was used in the Lass et al (1973) study and the stimuli in that of Lass and Wolden (1971) were all isolated vowel sounds. No real words were present in either experiment. This slim indication that musical stimuli might function in a similar way to verbal tokens can be viewed as a starting point for an investigation. However, it is clear that this finding in itself is not enough to make a serious case for a non-verbal analogue for a phenomenon which has been so firmly rooted in the field of speech perception.

The wide-spread acceptance of Warren's explanation of the VT effect had the consequence of reducing interest in the general phenomenon and particularly in the exploration of a possible non-verbal analogue. It was thought that although some illusory changes might occur with alternative types of stimuli, they did not constitute a genuine analogue, only a rather pale imitation of the verbal effect. This was very much led by the prevalent theoretical view that uniquely speech based mechanisms were the main factor in VT production. The essence of the general argument was that, although non-verbal stimuli could perhaps elicit some illusory changes, these were only peripheral and marginal whilst only speech could produce the dramatic transformational shifts which are the hallmark of the phenomenon. It is clear from this that a vital if implicit feature of this position lay in the quality of the new forms experienced by the subjects as well as quantity.

Such a line of reasoning is unanswerable while no experiments have directly compared speech and non-speech stimuli under the same listening conditions and within a coherent framework from which responses may be evaluated and equated.

Cross-Stimulus Comparison of Illusory Changes: A Framework

The fundamental difficulty confronting anyone seeking to construct a non-verbal analogue for the VT effect is the means by which illusory percepts elicited by differing types of stimuli may be equated. For example, does a pitch change in music equal a phonemic change in a word? More precisely, on what basis can the comparison be made?

The framework of analysis proposed here centres on the concept of 'stimulus identity'. In the occurrence of VT's the objective representation of the identity of the stimulus does not alter at all, it remains constant from the first presentation to the last. It is the subjective representation that suffers change. Thus, if coherent criteria for what constitutes a change in identity for each of the stimulus types could be clearly prescribed, then perceived transformations may be compared.

It may be argued that no comparison of stimulus identity can be made without addressing the concept of meaning. If Warren's explanation of the VT effect is correct, then the search for meaning is the underlying reason why the stimulus is re-arranged by the speech processing mechanisms. It is not denied that spoken language as our primary form of communication has a semantic subtlety, flexibility and range which cannot be equalled by other auditory stimuli. However, everyday sounds and music are not without their forms of meaning. The sounds of screeching car brakes and breaking glass are no less evocative than the words used to describe them. The actual noises can be more potent than their linguistic counterparts and music can convey a breadth and depth of emotional expression not available to other auditory forms. The symbolic resonances of certain instrumental parts or particular pieces of music should not be ignored.

The relative semantic merits of the possibly analogous stimuli may seem to be an interesting and problematic area for debate, but it is the contention here that it is essentially sterile, offering only a cul-de-sac in the understanding of the VT effect. The format of the stimuli used to elicit the phenomenon gives rise to this argument. The brief duration of each individual stimulus allows time for a spoken word or a phrase and would enable a short everyday sound to establish a simple sort of referential semantic (e.g. a breaking glass), but would not allow music to develop any emotional or symbolic meaning. Does this then preclude music from acting as an analogous auditory stimulus in producing illusory transformations? Perhaps so, and yet Lass, West and Taft (1973) have suggested, on the basis of their brief study, that music is the most likely analogue.

If the semantic nature of verbal utterances is assumed to be the engine driving illusory change, then there is an impasse both theoretically and practically. However, if the semantic aspect of the phenomenon is considered as one element of the stimulus, which may or may not transform, rather than as the cause of all the changes, then it is possible to proceed.

The reason for making this important assumption is that if the literature from Warren (1961a) onwards is examined, it can be seen that semantic changes in the identity of words are invariably accompanied by changes along other dimensions. An example of this might be that the stimulus "TRESS" is reported as changing to "DRESS" which, at one level, might be regarded as a change of meaning. But at another level, this may be seen as a change in the identity of the initial phoneme. So while the subject certainly does perceive a different word, it is not at all clear whether the psychoacoustic disruption results in an inevitable change in meaning, or whether a perceived change in meaning leads to a phonemic switch.

There is no evidence at the moment that pure semantic changes take place at all. It is true that subjects often report that they have "lost" the meaning of a word (i.e. the repeating stimulus is stripped of all semantic content and is perceived merely as a series of sounds), but no experiment has directly manipulated and measured pure shifts of meaning between, for instance, plausible alternatives in a set of words with more than one possible definition. Conversely, non-word stimuli have been shown to change into other non-words, demonstrating that purely phonemic transformations do take place.

It should be clearly understood that although there is no evidence that pure semantic changes take place, this line of reasoning does not deny the possibility that they could occur but seeks instead to emphasise the point that transformations in meaning are routinely linked with phonemic shifts. Hence, semantic changes should not be regarded as the "prime mover" in the VT process, but should be more realistically assessed as one of a number of characteristics possessed by a repeating token which might undergo transformation. This view is not, of course, compatible with Warren's theoretical explanation of the phenomenon. However, as has already been suggested, there are more plausible alternatives to Warren's view, and unless the role played by the semantic features of the stimulus is placed in a different perspective, there can be no logical starting point for a non-verbal analogue as no framework of response comparison can be constructed.

The basis for comparison which is advanced here arises from the type of illusory transformation which the subject perceives. The changes are classified in basically one of two ways; they either alter the identity of the original percept or they are peripheral. This concept can be understood more clearly by reference to an example. If the subject reports that the rate at which the stimulus is presented is increasing or decreasing, this certainly can be counted as a form of illusory change but may be regarded as peripheral to the identity of the token. In this situation, the subject claims to be perceiving the same word being spoken more quickly/slowly than the first presentation. This would be an example of a peripheral change. This may be contrasted with the type of illusion at the other end of the transformational spectrum where a subject might say that he/she is hearing a completely different word. In the latter case, the subject feels that he/she is experiencing a new item, and this would be classified as an identity change. Hence, if different types of stimuli produce comparable numbers of identity changes, then a compelling case for an analogue can be formulated. That is, if music or other auditory sounds produce a similar number of illusory transformations which can be considered vital to the integrity of the original percept, then it could be said that these stimuli are performing in a similar way to words.

At the top end of the identity continuum, the report that a musical motif had changed completely could be considered as equivalent to the perception of a new word. These gross and complete switches in identity can be characterised as global changes. There are, however, other changes which do not completely lose the original token but make sufficient inroads on its psychoacoustic integrity as to cause the illusion to be counted as a transformation of its identity. These are perceived changes which, although preserving part of the original stimulus, vary from it in some vital elements. An example of this would be a subject who is presented initially with the word "TRESS" and then reports it changing to the word "STRESS". Here, the original token is retained but a new and extra phoneme is added which changes the stimulus identity.This also applies to the situation where a subject loses a phoneme e.g. "TRESS" might become a non-word "RESS".

Neither of these two possibilities represent a global change in the percept as large sections of the original remain, but they do represent changes from the original identity of the token. These examples can be generalised to other stimulus forms. A subject might perceive new notes, or the loss of some of the original ones, during a trial with a musical phrase. It would be difficult to argue that this would have less of an effect on the musical identity of the motif than the alteration of a phoneme during a verbal trial.

It is therefore possible to generate a set of criteria for defining transformations which would constitute identity changes in different types of non- verbal stimulus material. Such a set was created and is listed in

Appendix 1. It embodies two basic categories of changes : identity changes and non-identity changes. The latter category contains transformations which do not affect the essential identity of the original percept. It is implicit in the formulation of this framework that if any form of stimulus is to be considered as a true analogue for the VT effect then it must produce a comparable number of identity changes as a word would produce under similar listening conditions. This tenet formed the starting point for the first experiment.

CHAPTER 2 Experimental Series 1

Experiment 1.i

Introduction

The general aim of this experiment was to examine the possibility of the existence of a non-verbal analogue of the VT effect. The historical and practical basis for a comparison has already been discussed so this brief introduction will centre on the theoretical implications of the possibility that auditory stimuli other than verbal tokens can be used to generate illusory transformations.

If an analogue can be satisfactorily demonstrated, then previous notions about the nature and function of the VT phenomenon will have to be re-examined. Central to this would be a reassessment of the idea that VT's are the outward sign of a mechanism uniquely dedicated to speech processing.

There are two corollaries of a successful challenge to the existing theory. First, the VT effect would have to be removed from the list of phenomena which purport to demonstrate the unique place which speech has in auditory perception. With the erosion of this evidence, the whole premise of the singularity of verbal stimuli would be brought into question.

Secondly, any models of the VT effect which characterise it as a speech based mechanism would need to be reassessed.

This experiment seeks to demonstrate a broad based and secure analogue in the non-verbal domain in order to promote these two strands of theoretical reasoning.

Method

Design

The experiment employed a single main factor - the nature of the stimuli used to elicit transformations - with three levels:-

1. Speech sounds - `Words'
2. Musical tokens - `Music'
3. Non-musical sounds - `Sounds'

These were used in a within-subjects design. This was chosen in preference to a between-subjects design in order to minimise the effect of the acknowledged large individual differences in susceptibility to the VT effect.

Three different measures of transformations were taken: time of first response from time of stimulus onset, nature of illusory changes and quantity of illusory changes. The purpose of using a multiplicity of measures was to ensure a sufficiently broad base of data which could provide converging evidence for a true analogue for the VT effect in non-verbal stimuli.

Stimulus selection, preparation and presentation

The criteria for the selection of stimuli in each of the three categories were familiarity and acoustic characteristics. In addition, stimuli in the `Words' condition were selected to include different parts of speech and a wide range of phonemes.

The stimuli used are listed in abbreviated form in Figure 2.1 below. Examples from each stimulus category can be heard on Tape 1.

Figure 2.1

	Words	Music	Sounds
1.	Daffodil	Abnormal Scale	Photocopier
2.	Cancer	Ba Ba Black Sheep	Printing
3.	Dungarees	Archers	Car Skid
4.	Circulate	Chimes	Lift Doors
5.	Minimise	Tristran	Static
6.	Aromatic	Bach	Camera
7.	Lilliputian	Fast Melody Line	Sonar
8.	Energetically	Three Notes	Jigsaw
9.	Drinking water	Rhythm	Crash
10.	Parliamentary debate	Fast Alto Line	Coin Drop

A Fairlight music computer in conjunction with traditional tape recorders was used to prepare the stimuli for presentation.

A single token of each stimulus, of not more than 2 seconds duration, was recorded and then repeated at a rate of 30 repetitions per minute. In this way, a 4 minute tape of each stimulus was constructed so that subjects could be presented with the experimental material in a convenient way, through headphones attached to a cassette recorder.

Subjects

The subjects were all undergraduates at the City University, London. Their ages ranged from 19 years to 36 years with an average of 24 years. The group of 13 consisted of male and female subjects. All subjects were paid for their participation.

Procedure

In order to accommodate all thirty stimulus presentations plus three practice trials, subjects were required to attend four separate test sessions each lasting one hour. Presentation of stimuli was counterbalanced across all test sessions in a Latin square. Before listening to any stimuli, subjects were given detailed verbal instructions which stated that they would hear a sound repeatedly played. They were clearly informed that the repeated stimulus would not vary but that a facet of the perceptual system might cause them to hear some change to the input. It was strongly emphasised to the subjects that a "no change" response was equally as experimentally interesting as the report of some change. Subjects were asked to respond to each stimulus by verbal report to the experimenter as soon as they had perceived any change, rather than waiting for the end of the trial.

Before the experimental trials began, subjects were asked to perform in three practice trials, one from each category, to ensure that they understood the experimental procedure. All responses made by the subjects during the stimulus presentation were recorded by the experimenter, as were any relevant comments made by subjects after each trial.

Response Recording

Verbal report was chosen in order to avoid any response bias which may have occurred should subjects have had difficulty in making a written response to musical and sound stimuli.

The transformations which are quoted in the Results and Analysis section may be more precisely regarded as new illusory forms of the stimulus percept. Hence, a subject who reported a single deviation from the original stimulus, even if that percept was heard repeatedly, would be recorded as having heard one transformation. Only when that altered to a new form would a second change be registered. It was decided that only new forms could be accurately recorded, as subjects engaged in describing an illusory percept could not be expected to simultaneously maintain a reliable count of the number of times any given change had occurred. Notes were kept of changes that were reported as repeating or alternating between different perceptual representations but these were not used as part of the final analysis. This protocol was used in all subsequent transformation experiments.

Results and Analysis

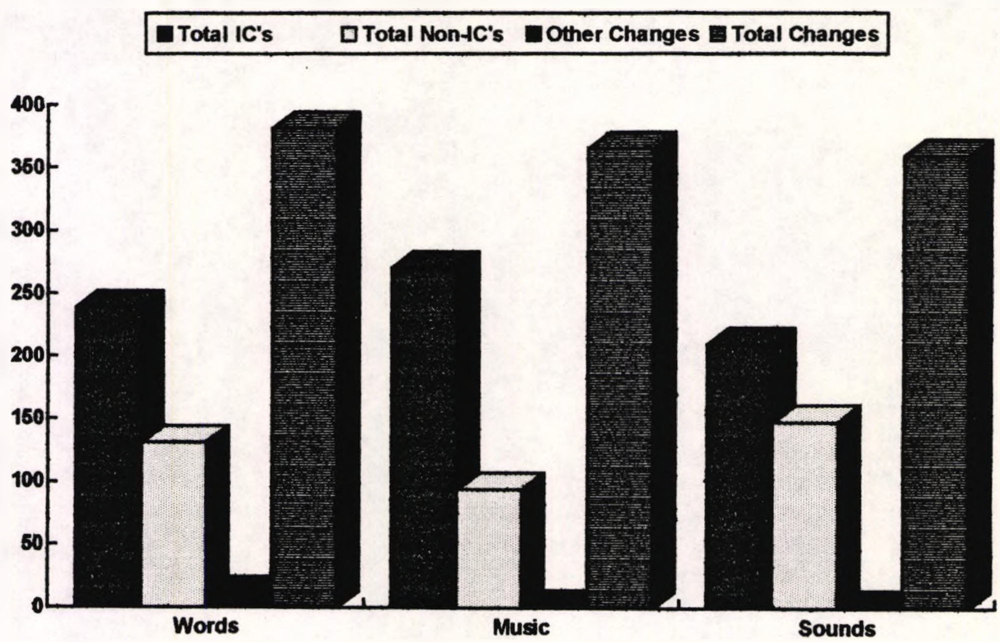
Table 2.1 summarises the data on illusory transformations in terms of Identity Changes (IC's), Non-Identity Changes (Non-IC's) and Other Changes. This data is also graphically displayed in Figure 2.2. A more detailed breakdown of transformations is given in Appendix 1.

Table 2.1

Illusory Transformations in Speech, Music and Everyday Sounds

	Words	**Music**	**Sounds**
Total IC's	239	270	210
Total Non-IC's	131	94	148
Other Changes	11	1	1
Total Changes	381	365	359

Figure 2.2 Total illusory changes in each stimulus category



It can be seen from the above that Words, Music and Sounds behave rather similarly in terms of eliciting transformations. The total number of changes appears very comparable across all categories.

Looking in more detail at the sub-categories of changes, the same pattern emerges. In each case identity changes exceed non-identity changes, although there are slight variations in the differences between these types of change as a function of stimulus category.

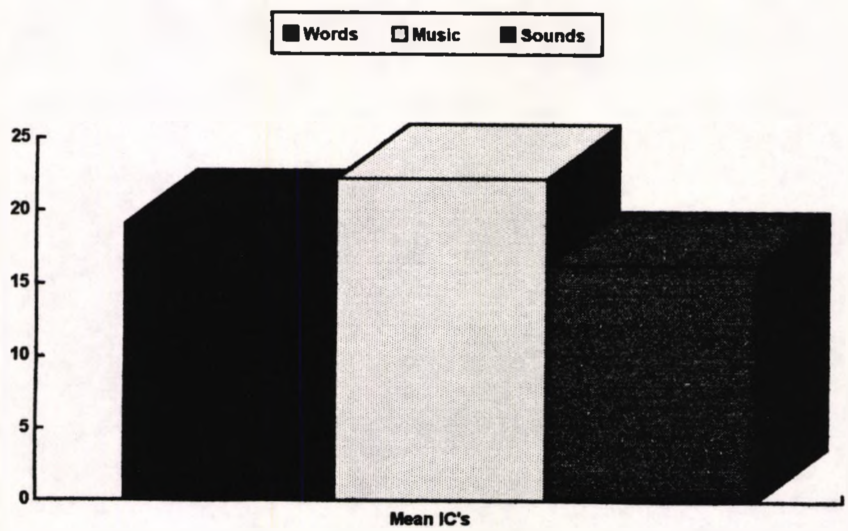
In order to make an appropriate assessment of the performance of Music and Sounds as potentially potent elicitors of illusory transformations, the analysis has concentrated on the critical section of responses which involve a gross distortion of the original stimulus, the identity changes (IC's). This is shown in Table 2.2 and graphically in Figure 2.3.

Table 2.2

Mean IC's for each stimulus category

	Mean IC's
Words	19.08
Music	22.25
Sounds	16.33

Figure 2.3 Mean Identity changes in each stimulus category



An ANOVA consisting of 1 factor with 3 within-subject levels was used to analyse this critical tranche of data. The outcome indicated a significant variance across the levels:

$$F(2,22)=4.19 \quad p<0.05$$

It had been predicted that if all three stimulus categories performed in a similar, and thus analogous, fashion, then no significant variation would be observed. However, the result shows a significant difference in performance across the categories. In order to examine this effect in more detail, a series of pairwise comparisons were carried out:

Words vs Music $F(1,22)=2.397$ $p>0.05$

Words vs Sounds $F(1,22)=1.808$ $p>0.05$

Music vs Sounds $F(1,22)=8.367$ $p<0.01$

A correction to allow for a increased probability of a Type 1 error due to repeated comparisons was introduced. This involved raising the required level of significance by the total number of comparisons - in this case, three. Thus, to be accepted at the 0.05 significance level, the comparison needs to reach a level that is more stringent by a factor of three, that is 0.017. Of the above comparisons, the only one to reach this level of significance is Music vs Sounds.

Music vs Speech & Sounds $F(1,11)=9.024$ $p<0.05$

Sounds vs Speech & Music $F(1,11)=11.132$ $p<0.01$

These comparisons and overall analysis demonstrate that, although the number of IC's elicited by Words does not differ significantly from those of Music or Sounds, the latter categories differ significantly from each other.

It is therefore reasonable to conclude that illusory transformations do take place beyond the domain of speech, as both sets of non-verbal stimuli induce levels of distortion similar to those produced by verbal stimuli.

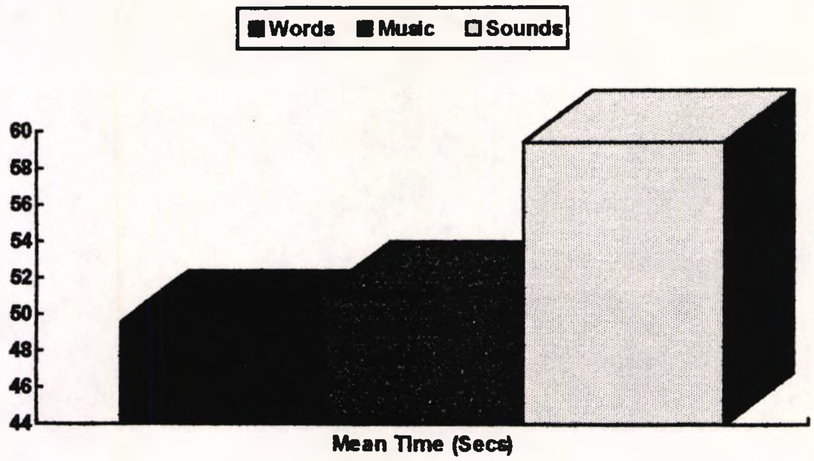
However, an unexpected aspect of the results is that Music and Sounds are different, suggesting the operation of mediating variables unsuspected pre-experimentally.

The other measure used to indicate potency of a stimulus to give rise to transformations was the mean time (and hence the number of stimulus repetitions) prior to the first reported illusory change of any kind. The data are shown in Table 2.3, and graphically in Figure 2.4.

Table 2.3 Mean Time to 1st Transformation

	Mean Time (Secs)
Words	49.57
Music	51.18
Sounds	59.39

Figure 2.4 Mean Time to 1st Transformation



The result of a one factor ANOVA with three within-subject levels, used to analyse these data, was:

$F(2,22)=5.823 \quad p<0.01$

This indicates that there is a significant difference between the stimulus categories in the time taken to report the first illusory transformation. Pairwise comparisons were performed in order to investigate the detail of this result :

Words vs Sounds $F(1,11)=7.592 \quad p<0.05$

Music vs Sounds $F(1,11)=5.301 \quad p<0.05$

These comparisons indicate that Words and Music both induce illusory transformations more quickly than Sounds. However, this measure should be treated with some caution, as subjects often experienced some difficulty in finding an appropriate verbal description for non-verbal transformations.

This applied particularly to the Sounds category, and may well be the source of some bias against an immediate report of the initial change in percept.

Discussion

A general overview of the results clearly shows that music and other sounds do produce transformations in comparable numbers to those elicited by words. But this alone could not be considered sufficient evidence to claim that a non-verbal analogue to the VT effect has been demonstrated. As discussed earlier, illusory changes reported by subjects do not always offer a comparison of like with like. It is the distribution between identity and non-identity changes which gives the best picture of the performance of the stimulus groups. That is, the critical feature is the number of identity changes in each category.

For example, it is possible to imagine a set of results whereby words and music produced a similar number of transformations, but the verbal stimuli elicited a large majority of changes in the perceived identity while the non-verbal stimuli produced non-identity transformations. In this situation, no analogue could be claimed as the words would clearly produce a more potent illusory effect, with music only eliciting peripheral changes in the percept. However, this scenario does not appear to be the case. In all three stimulus categories, changes to the identity of the percept were in the majority. This indicates that the pattern of performance was similar whichever stimulus category was used and clearly supports the notion of a non-verbal analogue for the VT effect.

There was a general expectation prior to the experiment that, even if music and other sound stimuli did perform in an analogous way to words, the verbal stimuli would still be the most potent elicitors of illusory forms. This did not prove to be the case. If the number of identity changes is used as the index of potency, then musical stimuli are the most effective. Identity changes are in fact the most convincing measure of potency. The reasons are twofold : first, subjects perceiving changes in the identity of the stimulus are experiencing the most gross distortions of the original percepts and, secondly, identity changes are accompanied by greater subjective confidence on the part of the subjects.

An important question posed by these data is why music should be such a potent elicitor of gross illusory distortions? This unexpected finding certainly could not be explained by reference to any of the speech based models of the VT effect as they would not predict any illusory changes from non-verbal stimuli. The most plausible explanation available is provided by the Associative theory advanced earlier as part of this thesis. This proposes that some tokens have internal representations which are connected in configurations which are more likely to produce illusory realisations when subjected to repeated stimulation. There is, at present, an insufficient database of experimental findings to enable a precise delineation of features which would make up the optimal configuration for producing illusory changes. However, a different analysis of these results seems to point to some of the more important axes along which this optimal configuration can be assessed.

On one level, the pairwise comparisons made between Words and Music and between Words and Speech reveal no significant differences and clearly support the notion of a non-verbal analogue for the VT effect. However, with further comparisons another pattern emerges. When Music is compared with a combination of Words and Sounds it elicits significantly more changes in stimulus identity ($F(1,11)=9.024$ $p<0.05$) than the other categories. Conversely, when the category of Sounds is compared with the others in the same way, it produces significantly fewer identity transformations ($F(1,11)=11.132$ $p<0.01$). Furthermore, when Music is compared directly with Sounds, a significant difference emerges.

This appears to indicate that Music and Sounds have more in common with Words than with each other. This pattern makes some sense if the similarities and differences between the types of stimuli are examined. The first point to note is that Words and Sounds share a semantic element in this experimental context, if only in a directly referential sense. The sound of the lift doors closing conveys the same information as the sentence "the lift doors are closing". This similarity does not, of course, extend to abstract concepts, but where sounds refer to real world events, they convey just as much semantic force as the words which may be used in descriptions. Music has often been used to symbolise elements of the real world, but in the short (two seconds) bursts used here, this feature of more extended musicality may be set to one side for the time being. The fact that this element of a referential semantic is shared by the two categories which proved to be the least potent elicitors of changes in stimulus identity, suggests that "meaning" is not a necessary feature in order for transformations to occur. Indeed, this is borne out by past research which has shown that non-words can produce illusory transformations.

A second, more fruitful, concept might be that both Words and Music can be said to have, in a broad sense, a grammar. The idea here is that the elements which go to make up the stimuli can only combine in certain ways. The limitations codified by a grammatical structure may make stimuli from these categories behave in a regular and coherent fashion. That is, when a stimulus enters such a system, it has the effect of not only priming the representations that are likely to follow, but also of inhibiting the grammatically nonsensical options and thus speeding up the process of normal recognition by limiting the number of prioritised connections in the system. A more in-depth discussion of the role of grammar in music is provided by Lerdahl and Jackendoff (1983).

The general point can be seen more clearly in relation to Sounds. If, for instance, a person enters a large railway station, his/her perceptual system would be primed to receive a very broad range of aural input, from the sound of the ticket machines and trains to the chimes which precede announcements. This requires a system to be ready for a diffuse and surprising range of stimuli. In contrast, a stimulus which conforms to a grammatical pattern primes only a restricted number of options. Hence, tokens which stimulate a limited number of other representations within an existing structure tend to be the most potent elicitors of gross illusory transformations.

This analysis, although post-hoc, provides a coherent explanation of the results where the meaning of a stimulus seems to be a less cogent factor in eliciting identity changes than the form in which a stimulus is related to other representations. It appears that fewer stronger connections are more likely to produce illusory effects. The rationale behind this is that if a restricted number of representations are primed to a higher level of readiness, they are more likely to fire on repeated presentation of the original stimulus than a more diffuse group. This line of reasoning can be followed through the other two experiments on this topic.

Apparent confirmation that Words and Music are more potent elicitors of transformations than Sounds comes from the time taken by subjects to report their first transformation. Here, both Music and Words were significantly faster than Sounds, but were very similar to each other. As a general rule of thumb, this measure is worth reporting, but a note of caution should be added that subjects often found it more difficult to verbalise their experience in the Music and Sounds categories. This may well have led to a brief delay in reporting their first illusory percepts, so the times recorded for these stimuli should perhaps be regarded as slight, but consistent, over-estimates. The individual differences exhibited in this regard complicate matters to the extent that adding a correction factor becomes impossible.

The final area that requires consideration is the comments made by the subjects during and after the trials. This raises at an anecdotal level the differential potency of the stimulus categories. When subjects had experienced a particularly gross distortion of the original stimulus, a common reaction was to refuse to believe that the tapes only contained repetitions of the original tokens. In fact, in debriefing sessions, some subjects asked and were allowed to 'browse' aurally through the tapes (by stopping and starting them at different points) in order to try to find the illusory percepts which they believed to be present on the recording. The Music stimuli generated far more disbelief reactions than did the other categories. This does not imply that Sounds and Words did not elicit strong transformations, but rather that subjects seemed more willing to accept the occurrence of distortions, especially when words were the repeated token. Also, the more the illusory percept differed from the original input, the more the subjects were inclined to believe that the tape had been physically changed by the experimenter. This gives a rough and ready barometer of the depth of conviction of subjects that the stimulus had actually changed. The notion of an analogue is fully supported by this anecdotal testimony as convincing changes were experienced by subjects in all categories.

The search for non-verbal parallels to the VT effect is continued in the next experiment.

Experiment 1.ii

Introduction

The purpose of this experiment was to replicate and extend Obusek and Warren's (1973) study using musical stimuli. Their paradigm consisted of removing a phoneme from a spoken word, replacing the missing section with white noise, then comparing this new stimulus with the performance of the original token under VT conditions. This, of course, is directly related to the most clearly formulated theoretical explanation of the VT effect advanced by Warren (1982). He proposed that the way the transformations tended to cluster around the missing segment of the word "MAGISTRATE" demonstrated that phonemic restorations and VT's were both active speech processing mechanisms acting to reconstruct the flawed input in order to create an intact percept. As previously discussed, it is the contention here that these two phenomena are essentially different in nature and function.

If music were to behave in a similar way to verbal stimuli under this paradigm, this finding would primarily undermine the central tenet of Warren's position by dislocating the combination of VT's and PR from a speech-specific interpretation. The extension of the present study to include an assessment of other variables and further forms of subsidiary measurement was intended to shed some light on the potential interaction between the two phenomena.

Method

Design

The design followed that of Obusek and Warren (1973), using musical themes as stimuli. The main variable was the degree of intactness of the musical themes. Stimuli were presented either : Intact (Control condition), or with a small intra-stimulus segment - a single note - removed and replaced with white noise, Deleted (Experimental condition). The object of this manipulation was to allow a comparison of illusory transformations occurring in the immediate vicinity of the noise burst with those occurring in the same section of the intact stimulus.

In order to extend this study to provide additional information about the behaviour of music as a stabilised auditory input, two other subsidiary variables were studied :-

1.Familiarity. Half of the themes used were very widely recognised classical themes, whilst the other half were extremely obscure.

2.Metrical position of the missing note. Not every note is of equal value in the metrical structure of a piece of music. In short, some notes are more important than others. In half of the tunes, a note from a strong metrical position was removed e.g. first note of a bar. In the other half, notes were removed from weaker metrical positions in the theme.

Thus the design involved three variables, each with two levels : Intactness of themes (Intact - Deleted), Familiarity of theme (Familiar - Unfamiliar) and Metrical strength of position of deleted note (High - Low). This formed a two by two by two within subjects design, as shown below.

Intact				Deleted			
Familiar		Unfamiliar		Familiar		Unfamiliar	
High	Low	High	Low	High	Low	High	Low
1	2	3	4	5	6	7	8

In order to glean as much information as possible about the subject's perceptual experience, a rating was obtained from each subject indicating how certain they were that the music continued during the noise bursts.

Stimulus Construction, Preparation and Presentation

Eight musical themes were used, four familiar and four unfamiliar, all written by the same composer. All subjects who took part were familiar with the four well known themes but did not recognise the others. A full list of these themes is given in Appendix 1. Below (Figure 2.5) is an example of a theme which was used. The Figure shows the segment which was removed and replaced with a white noise burst. In the Intact condition, the theme was heard in full.

Figure 2.5

Prokofieff, Sonata in D, Op 94,

4th movement, 3rd theme.

* Denotes missing/replaced note

Musical stimuli were generated and edited using a Casio C Z 1 keyboard and an Atari ST 1040 computer with a Sternberg Pro 24 Version 1.1. programme. All stimuli were repeated at a rate of thirty repetitions per minute over a four minute period. All subjects heard the stimuli through earphones connected to a standard Sony TC - FX3 cassette player.

Subjects

Subjects were all undergraduates at the City University, London. Their ages ranged from 19 years to 33 years with an average of 23 years. The group of 13 consisted of male and female subjects. All subjects were paid for their participation. None of these subjects had taken part in the previous experiment.

Procedure

Two test sessions were required. Each session lasted one hour. Subjects were given explicit instructions to report any changes in the stimulus immediately they were heard to the experimenter. They were also encouraged, as far as possible, to describe accurately the position of any illusory change they might perceive. Before the experiment began, subjects were given a practice trial to acclimatise them to the testing environment and to check their comprehension of the verbal instructions.

Presentation of stimuli was counterbalanced across all test sessions in a Latin square. Transformations were recorded under the same protocol as Experiment 1.i

After each stimulus trial, subjects were asked if they were:-

(a) Familiar with the theme they had heard.

(b) In the case of the Experimental condition, sure that the music had continued during the noise burst. This judgement was made on a 7-point rating scale, where 1 indicated certainty that the music had not continued during the extraneous noise burst, and 7 indicated certainty that the music did continue during the noise.(see Appendix 1).

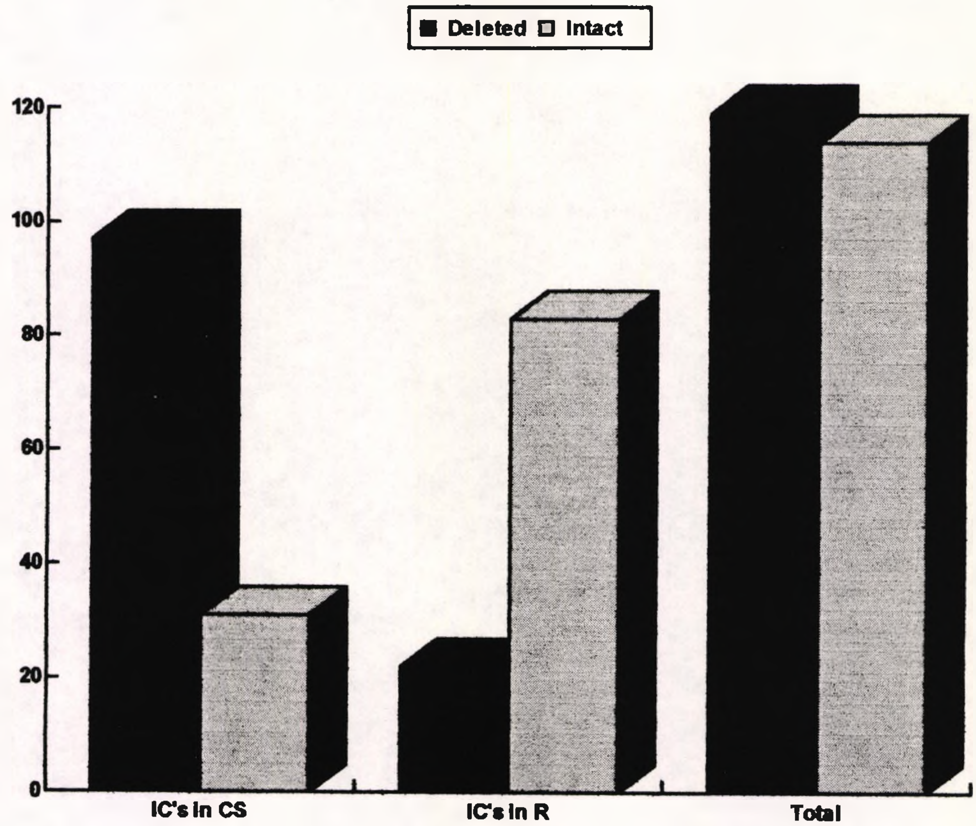
Results and Analysis

The data considered here consist of the number of illusory transformations reported by all subjects. Table 2.4 represents the illusory changes (IC's) reported at different locations in the themes. Possible locations were divided into two categories : one, the critical segment (CS), was defined as encompassing the replaced/intact note plus two notes on either side, and the other, the Remainder of the Theme (R). This is graphically represented in Figure 2.6.

Table 2.4 Perceived Location of Illusory Changes

	IC's in CS	**IC's in R**	**Total**
Deleted	97	22	119
Intact	31	83	114

Figure 2.6 Perceived Location of Illusory Changes



The difference between the total number of changes in the two conditions is minimal. This is confirmed by the results of an ANOVA performed on these data. This clearly indicates that there is no overall difference in the total number of changes reported for the Intact and Deleted conditions:

$F(1,12)=0.228$ $p>0.05$.

However, the distribution of the transformations is significantly different. This can be shown through the significant interaction between the location of illusory changes and the condition in which subjects were listening:

$F(1,12)=24.193$ $p<0.0004$.

This is shown above in Figure 2.6. It is apparent that in the Deleted condition the transformations displayed a significant tendency to cluster around the note which had been replaced with white noise - that is, in the Critical Segment. This analysis is supported by two other relevant pairwise comparisons:

CS (Deleted) vs R (Deleted) $F(1,12)=17.24$ $p<0.01$

CS (Deleted) vs CS (Intact) $F(1,12)=13.053$ $p<0.01$

These comparisons demonstrate that the critical segments of the stimuli in the Deleted condition elicited significantly more illusory changes than the remainder of the themes. Also, the critical sections close to the replaced notes induced significantly more transformations than the corresponding segments in the Intact condition. These findings lead to the conclusion that replacing a missing note with white noise has a strong effect on the way that the perceptual system produces illusory changes from a given token under conditions of stabilised auditory input.

Other findings to emerge from the manipulation of the other variables in the experiment are perhaps of lesser relevance to the theoretical context of the study, but are nevertheless of some interest.

The effect of familiarity of a theme is shown in Table 2.5 and Figure 2.7

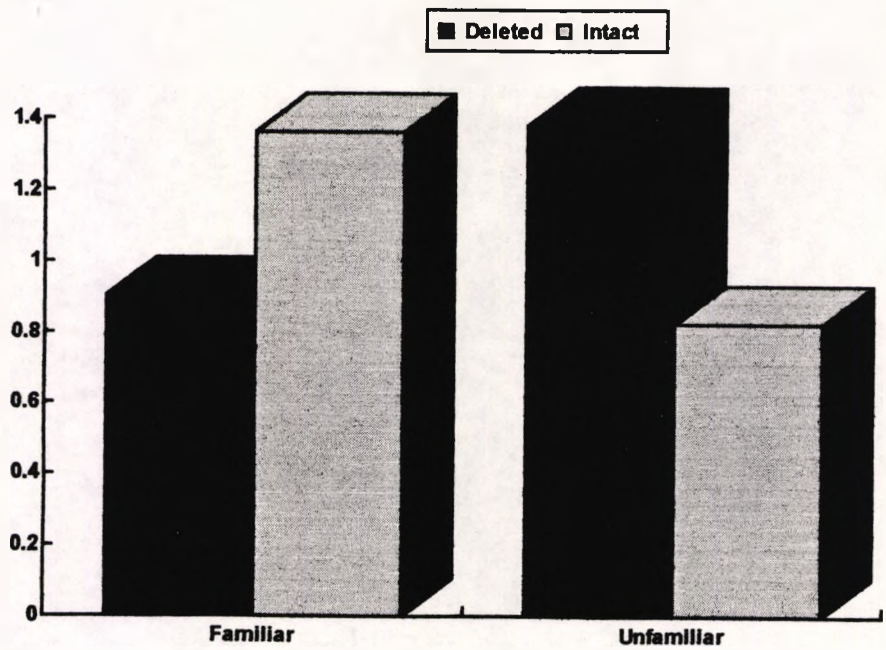
Table 2.5

Mean Number of Illusory Changes for Familiar and Unfamiliar Stimuli.

	Familiar	**Unfamiliar**
Deleted	0.9038	1.3846
Intact	1.3654	0.8269

Figure 2.7

Mean Number of Illusory Changes for Familiar and Unfamiliar Stimuli.



The interaction between the familiarity of the stimulus and the condition in which it was heard gave rise to the following value :

$F(1,12)=8.339 \quad p<0.05$

Thus while there was no difference in the overall number of transformations occurring in familiar and unfamiliar themes, there was a significant interaction between familiarity and whether there was a deleted note or not. More transformations occurred in the Deleted condition when the theme was unfamiliar, while more transformations occurred in the Intact condition when the theme was familiar.

Another effect involved the relative metrical importance of the missing/replaced note. Although there was no overall effect, more illusory changes were reported in the critical segments of themes where the missing note occupied a position of higher metrical value. This was revealed by a pairwise comparison between the number of transformations occurring near replaced notes of high and low metrical value in the Deleted condition:

$$F(1,12)=9.432 \quad p<0.01$$

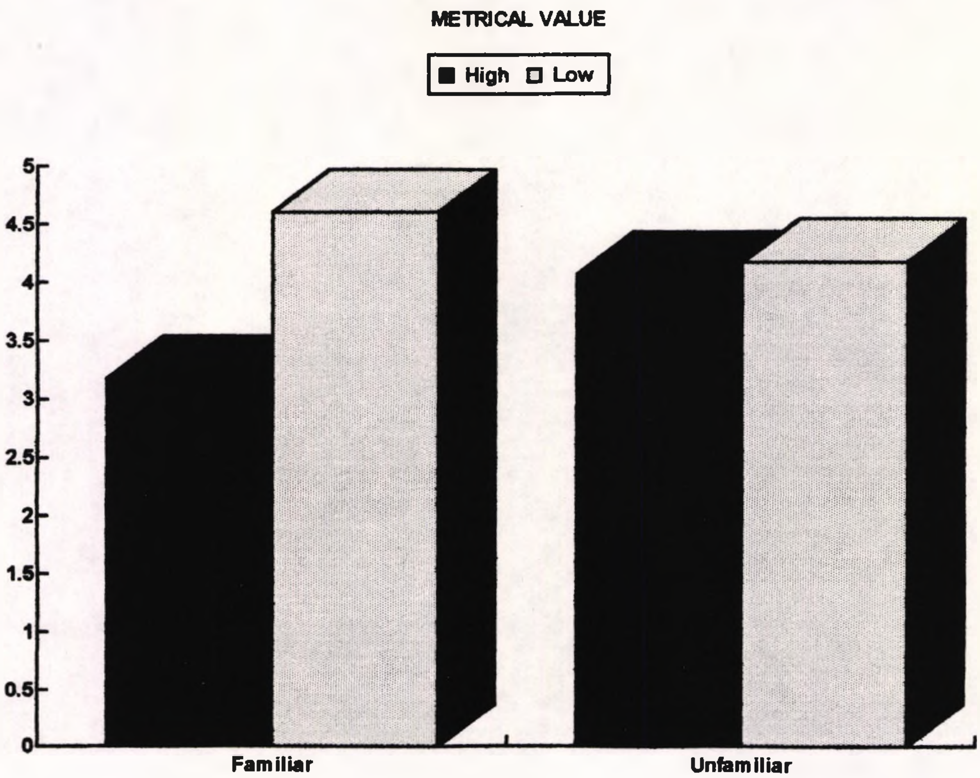
Together with a record of the number of illusory changes, a measure of the level of perceived continuity of themes in the Deleted condition was taken in order to assess the degree of restoration for missing/replaced notes while listening to repeated presentations of a stimulus. As previously outlined, this was measured using a rating scale of 1: Did not continue, through to 7: Certain of continuity. This is shown in Table 2.6 and graphically in Figure 2.8.

Table 2.6

Mean Continuity Ratings for Missing Notes in Repeating Musical Themes

Metrical Value	Familiar	Unfamiliar
High	3.176	4.071
Low	4.607	4.179

Figure 2.8



An ANOVA performed on the rating data revealed no significant difference between the ratings of continuity given to Familiar and Unfamiliar themes:

$$F(1,13)=0.611 \quad p>0.05$$

However, there was a significant difference between the restoration levels of missing/replaced notes of different metrical value. Notes of a low value induced more restoration than those of a higher value:

$$F(1,13)=5.861 \quad p<0.05.$$

These findings do not directly relate to the main theoretical context of this topic, they are, however, highly relevant to later discussions of the contiuity effect. These discussions provide a more in depth exegesis of the role of metrical structure in the restoration of missing notes.

Discussion

The first version of this experimental paradigm was designed by Obusek and Warren (1973) in order to establish a direct and explicit link between the speech based re-organising processes, which were the proposed cause of the VT effect, and the phonemic restoration effect. The findings of Experiment 1.i demonstrated that strong and compulsive illusory changes occur beyond the confines of verbal stimuli. Hence, the cornerstone of any purely speech based explanation has been removed.

However, what remains of the Obusek and Warren formulation is the connection it forged between two previously disparate phenomena - transformation and restoration - since VT's clustered in significant numbers around the missing phoneme. This connection is robustly confirmed by the findings of Experiment 1.ii, where the distribution of illusory transformations is distorted in just the same way by the presence of a missing/replaced note. Obusek and Warren's explanation of this effect derives from the view that both phenomena are generated by the same mechanisms. However, there is a logical problem at the centre of an argument which asserts the fraternity of two phenomena which seem to have opposite effects on a percept, transformations being essentially disruptive and restorations functioning only to preserve meaning and continuity. This proposition could be sustained while the speech based element of the theory restricted the nature of the processes involved. But to maintain the argument in the light of the new evidence would be to assert that the same processes are involved in preserving the contextual meaning of a word in a sentence as those which dissolve a musical theme into a mad jumble of notes. This inevitable conclusion seems to stretch the theory beyond the point at which it makes any explanatory sense in terms of its original tenet that transformations and restorations are manifestations of the same speech processing mechanism.

It is the contention here that a more valid interpretation of the findings of Obusek and Warren and Experiment 1.ii. is that the effect observed in both studies is caused by an interaction between two separate and generically different auditory phenomena. As suggested earlier, illusory transformations associated with repeating stimuli should be regarded as percepts which are generated only through the particular nature of that experimental paradigm, while restoration of a missing segment of auditory .information clearly reflects an adaptive process which maintains the sense of a token in relation to its context throughout brief interruptions.

The question which remains is, why do these two phenomena interact at all? An attempt to answer this question is provided in the General Discussion at the end of this section when all the evidence has been presented.

Experiment 1.iii.

Introduction

The aim of this study was to investigate whether a purely cognitive level of priming can have any effect on the perception of illusory changes when complex (non-verbal) sounds are used as repeated stimuli. The underlying rationale is that, if more transformations are elicited under this form of priming, then it may be inferred that the cause relates to an increase in activity amongst a critical set of interrelated representations, thus facilitating the spread of associative stimulation generated by the persistent nature of the stimulus.

Thus this experiment can be viewed as a direct test of the Associative theory's explanation of why the perceptual system produces illusory changes under conditions of repetitive stimulation. This model views the cause of all transformations as the stimulation directed to related representations of the repeating stimulus. Hence variables which influence the interconnection of internal representations should affect the perception of illusory transformations.

In the same way that an overheard remark about the 'seaside' may cause the listener to think about his/her holiday, it might be that a prime consisting of only high level descriptive information can significantly enhance this illusory phenomenon by opening up pathways to related representations.

The facilitation of transformations under such a paradigm could only be explained in terms of the Associative theory (A.T).

Method

Design

The experiment explored the effects of priming on the perception of transformations during repetitive auditory stimulation. The prime consisted of a description of the stimulus given prior to the listening task. In the Experimental condition subjects were given a general, yet colourful, description of each stimulus. In the Control condition no description at all was given. Thus, a between-subjects design with two levels of a single variable was employed.

The purpose of the descriptions was to act as a non-specific prime for a network of potentially related representations. The type of prime, a verbal description, as opposed to another complex sound, was chosen deliberately to contrast with the stimulus in order to examine the possibility that raising the level of expectation for a class of sounds need not be limited to a presentation of a generically similar token. This would show that a repeating stimulus activates associated representations at various different levels within the system.

Stimulus Construction, Preparation and Presentation

The stimuli were complex sounds (non-verbal and non-musical). They consisted of a diverse selection of items ranging from animal cries to machine noises. See Appendix 1 and the Taped examples.
The sounds were sampled using a Korg DSS 1 and were recorded on to continuous cassette tape at a rate of 30 repetitions per minute. This method of construction allowed a conventional Sony TC- FX3 cassette recorder to be used to present the stimuli over headphones.

Subjects

Subjects were all undergraduates of the City University, London. Their ages ranged from 19 years to 28 years with an average of 22 years. The group of 19 consisted of male and female subjects. All subjects were paid for their participation. None of these subjects had taken part in the previous experiments.

Procedure

Each subject was initially given a set of instructions which informed him/her that they would be listening to some repeating stimuli and that their task would be to report any changes they perceived in those stimuli. Subjects were then given a practice trial, to ensure that they understood the instructions. Subjects in the Experimental group then received a detailed verbal description of each sound they were about to hear. The sound was then played through the subjects' headphones for 120 repetitions (four minutes). Control group subjects heard the same sounds repeated the same number of times but did not receive the stimulus description. Subjects were randomly allocated to Control or Experimental groups. The order of stimulus presentation was randomised across subjects.

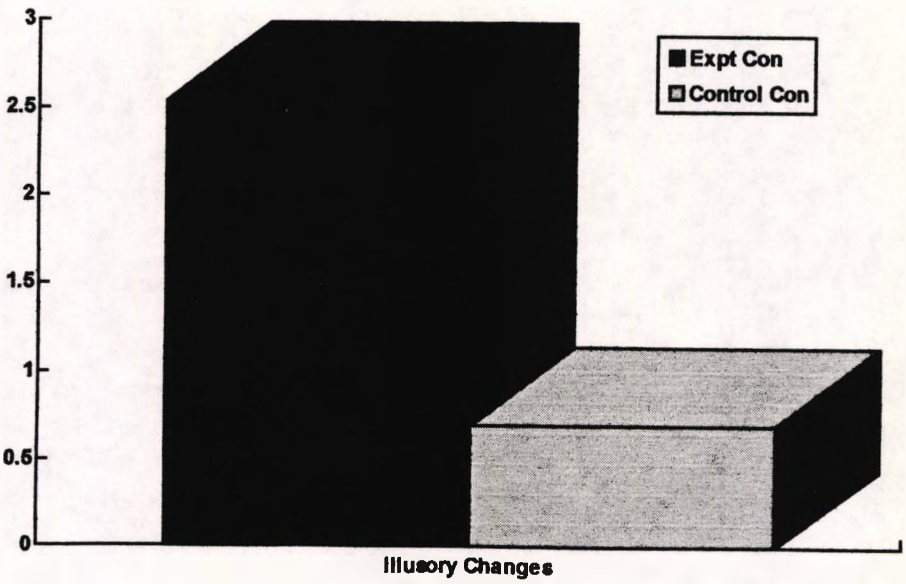
Results and Analysis

The results shown in Table 2.7 and graphically in Figure 2.9 illustrate the mean number of illusory changes per stimulus trial in each condition.

Table 2.7

	Illusory Changes
Expt Con	2.540
Control Con	0.700

Figure 2.9



The analysis used was a between-subjects one way ANOVA which revealed a significant difference between the number of transformations reported in the Experimental and Control conditions:

$F(1,18)=12.986 \quad p<0.01$

This clearly demonstrates that subjects who were given a prior description of the stimulus were prone to a far greater number of illusory changes than those in the Control group. Hence, the conclusion to be drawn is that a cognitive-level prime can significantly increase transformations.

Discussion

The result of this experiment indicates strong support for the Associative theory of illusory transformations in that a high level of cognitive priming, which was not specific to any auditory configuration, caused a highly significant increase in the reported number of illusory changes. Warren's model of the VT effect certainly cannot account for these results. The only adequate explanation is that the cognitive priming which subjects received facilitated the associative process which produces illusory changes. That is, repeated presentation of the stimulus was much more likely to activate associated representations if the related area had already been stimulated through descriptive verbal priming.

CHAPTER 3

General Discussion - Experimental Series 1

It is clear from the evidence presented in this section that none of the theories which have been advanced to date adequately explains illusory transformations in the perception of repeating auditory input. Warren's model would explicitly exclude non-speech illusions but, since a strong analogue has clearly been demonstrated, this position must be regarded as untenable. The primary purpose of this discussion will be to show how both recent and previous results can fit into the new Associative theory explanation of this phenomenon.

The central tenet of this theory is that the illusory percepts which subjects experience are not the product of a normally functioning system, but are caused by its persistent subversion. When any sound or word is perceived, associated representations are also stimulated. Normally, this adaptive process allows a more rapid response to new input but, when the original token is continuously repeated, the ongoing stimulation of related items ceases to fulfil its natural function and fires illusory elaborations of the real auditory pattern. Hence, the transformations are provoked by the nature of the stimulus undermining the normal operation of the perceptual system.

This theory would also emphasise that the perceptual system adaptively facilitates the processing of auditory domains which have a formal structure by priming related representations and segmenting input according to grammatical rules. Hence, in domains where stronger connections occur between structurally plausible items, it is more likely that the spread of inappropriate stimulation will also be enhanced when normal input conditions are subverted. This position is supported by the results from Experiment 1.i where Music and Words proved to be the most potent elicitors of illusory changes. These domains share the common element of grammatical structure whereas the category of Sounds, which does not lend itself to any syntactical descriptions, was shown to be least effective in producing transformations. This does not imply that stimulus domains which do not conform to any grammatical rules are not subject to associative illusory stimulation, but that the process is enhanced by connections which facilitate a structurally appropriate percept.

To explain this notion of grammatical facilitation further, it should be emphasised that this view of illusory transformations theorises that all areas connected to the repeating item - for example, phonemic, syntactic and semantic areas in the case of a word - receive associated stimulation. All of these areas are reflected in the responses made by the subjects, but domains which have systems which reflect the structural nature of the input, function adaptively to anticipate the next stimulus. It seems that this normally adaptive activity only serves to speed the spread of activation to related representations.

While the Associative theory would predict the superior performance of stimuli which relate to existing syntactic constraints, it would not foresee that there might exist an optimum set size of related representation. This is what seems to have happened in Experiment 1.i where Music elicited significantly more illusory changes in identity per token than a combination of the other two categories. The implication of this result is that the interconnection of musical representations offers an optimal environment for the spread of associated stimulation. This may be because there are simply fewer interconnected representations that are activated by any one repeating token, so that each receives correspondingly more stimulation, and that these connections are more proactive in the spread of associative stimulation. Conversely, there may be strong semantic restraints on the emergence of certain associated representations in speech and other everyday sounds.

Both of these options are purely speculative at this stage as there are no other data which can be called upon to resolve the matter. However, at a more concrete level of analysis, the results from the other two experiments in this section also confirm the findings of Experiment 1.i and conform to the precepts of the Associative theory. Both studies demonstrate that illusory changes can be elicited from non-speech stimuli, re-enforcing the primary theoretical thrust of this thesis. This being the case, none of the theoretical explanations in the literature is adequate. The only model proposed to date which can explain the findings is the model outlined here. Any notions predicated on the uniqueness of speech perception are clearly of no use given the findings. Experiment 1.ii shows this point clearly. Obusek and Warren (1973), who first used this paradigm, claimed that it established the link between transformational and restorative speech processes. However, the basic premise is undermined if verbal stimuli are not the only forms which can elicit illusory percepts. Also, the Obusek and

Warren construct carried with it an inherent logical flaw : if these two mechanisms are intended to work in unison to create the final percept, then why do they appear to have opposite functional effects? That is, when restoration is observed, missing segments are reinstated in order to preserve the meaning of the token within its context, whilst the very definition of a transformation is that it has distorted the original token, often beyond recognition. The new model offers a set of ideas which are much more consonant with the evidence since, within its framework, the two phenomena are regarded as wholly separate. Restoration reflects the adaptive normal functioning of the system enhancing disrupted input, whilst illusory transformations are seen as the product of a system pressured into dysfunction. The results obtained by Obusek and Warren (1973) and here in Experiment 1.ii can be interpreted in a new light as two disparate phenomena observed interacting rather than as similar processes acting in unison.

However, the core of the above interpretation rests on the exact nature of the interaction. It would be useful at this point to reflect on the central aspect of these studies. Both experiments found that when a small section of a repeating stimulus was removed and replaced with an extraneous noise, the illusory transformations reported by subjects gathered to a disproportionate degree around the missing segment. This proved to be the case for speech stimuli (Obusek and Warren (1973)) and music (Experiment 1.ii). Given that the effect is genuine (i.e. that where restoration processes are invoked, illusory transformations are multiplied) and given the premises that purely speech based explanations are invalid and that the two phenomena are generically different, what is the solution?

The root of the solution stems from the way in which the perceptual system functions in order to restore degraded or missing parts of the stimulus. A more fully elaborated discussion of how restorative processes may work is undertaken later in the light of new evidence but, for the moment, a basic description must be used so that this line of reasoning can be advanced.

The basic premise is that in order to reinstate a missing segment which will fit an initially ambiguous context, a number of plausible alternative representations must be stimulated to a state very close to conscious realisation for the subject to achieve the sensation of a complete percept when the disambiguating evidence occurs.

For example the word "Convergence" might be used as a stimulus with the phoneme **/v/** removed and replaced with a noise burst.

Figure 3.1

Noise Onset (N1) ➔**/** **/** ← Noise Offset (N2)

CON **/ V /** ERGENCE

Figure 3.1 shows that it is not until N2 that the perceptual system is able to disambiguate the stimulus. In fact, at N1 the word could end in a number of different ways, for example: Con***f***use, Con***t***empt, Con***c***ave, each requiring a different phoneme at the critical point to complete the token. The perceptual system must be prepared for the word to end in a number of different ways and it can be presumed that this is accomplished by stimulating a range of related representations. The restorative process is then enhanced by having the appropriate section ready for integration.

However, this system, which is designed to reconstruct imperfect input on a single presentation, acts as a catalyst for the spread of associated stimulation which is generated by repeated presentations of the same stimulus. Hence, the findings of Obusek and Warren (1973) and of Experiment 1.2, where transformations cluster around the area of the missing speech sound/musical note, are the result of the higher than normal priming activity generated by the system in order to cope with a specific area of uncertainty within the repeating token. Verbal reports from subjects in Experiment 1.ii indicated that not only was the distribution of the transformations radically altered, but the types of illusions experienced were related to the missing segment. The adaptive process of restoration effectively increases the likelihood that illusory percepts will occur in the vicinity of the missing segment of an incomplete stimulus. Reconstructive mechanisms provide an extremely fertile environment in which transformations can multiply.

The underlying idea proposed here is that an increase in stimulation of a specific area facilitates the occurrence of transformations. Converging evidence for this contention is provided in Experiment 1.iii. In this study, a set of unusual (non-musical, non-speech) sounds were used as stabilised auditory stimuli. One group of subjects was primed by receiving a brief description of the sounds, the other group was given no information. The result clearly demonstrated that the subjects primed prior to each trial perceived significantly more transformations than the others. Hence, it would seem reasonable to suppose that the extra stimulation imparted by simply describing the tokens facilitated the spread of associative activation among related representations. This adds credence to the notion that transformations can be closely linked to an increase in activity in a prescribed area, whether this is instigated by a piece of information exerting purely top-down priming influence, or by the automatic processes involved in restoring a missing section. That is, if connections between representations are at a heightened level of readiness, then the associative stimulation generated by a repeating stimulus is more likely to result in some of the interrelated representations rising to a level of conscious awareness.

In conclusion, the character of this phenomenon must be viewed in a completely new light. The old orthodoxy which portrayed verbal transformations as one of the uniquely creative facets of speech processing must be put to one side along with the functional explanations which supported those ideas. The evidence does not tally with the previously accepted explanations but it does fit in with the theory of illusory transformations advanced in these pages. This, of course, does not mean that the new set of ideas is correct, but it does present a challenge to anyone who wishes to study this effect in the future. The content of this challenge is summarised by the points listed below:-

1. Illusions produced by stabilised auditory input are not restricted to speech.

2. The illusions are not the result of normal processes but are elicited by the subversion of those processes.

3. The direct cause of the illusions is the excessive stimulation incurred by related representations when a token is repeatedly accessed.

4. Transformations are facilitated within stimulus groups which have grammatical/structural connections.

5. The connection environment in which associative stimulation can spread is influenced by involving normal processes (e.g. priming) which adaptively stimulate related representations.

PART TWO

THE CONTINUITY ILLUSION

The phenomenon of illusory auditory continuity occurs if a deleted portion of a sound appears to carry on through a louder noise which has in fact replaced part of the original stimulus portion. The first report of this illusion seems to be by Miller and Licklider (1950), who used both speech and non-speech sounds as stimuli.

However, after their research this topic divided into two separate branches; non-speech (mostly concerned with pure tones) and speech. Research on the non-speech aspects of the illusion have made fairly regular appearances in the literature since Miller and Licklider's study, but interest in the speech-based implications of the phenomenon did not reappear until Warren (1970) christened the verbal version "phonemic restoration". Although often quoted, this aspect of the illusion has not generated a large body of research. As Samuel (1981) put it "phonemic restoration has been very widely cited and very little studied". The present review of past studies will initially treat the two branches of research into the illusion of continuity as distinct.

CHAPTER 4

Phonemic Restoration

The first evidence for this effect came from a study by Miller and Licklider (1950). It involved recordings of monosyllabic words which were interrupted by silent gaps 10-15 times per second such that the on-time and off-time were equal. The effect was that the words became unintelligible. Miller and Licklider then filled the silent gaps in their stimulus words with a broad band noise which was slightly louder than the words. This failed to increase the intelligibility of the words although the stimuli sounded more "natural". They described their effect in visual terms saying that the sort of continuity that the subjects experienced was like viewing a landscape through a picket fence, "the pickets interrupt the view at regular intervals but the landscape is perceived as continuing behind the pickets".

It seems to have been Warren (1970) who first coined the phrase "phonemic restoration". Unlike Miller and Licklider (1950), Warren removed only one phoneme from the word "legislatures" (the first /s/). This he replaced with a louder extraneous noise and presented the word as part of a recorded sentence. Subjects perceived the whole word complete with the missing phoneme and were unable to locate the extraneous noise, hence, the "phonemic restoration effect", as subjects perceptually reinstated the missing segment. This finding was subsequently confirmed and expanded by Warren and Obusek (1971).

Warren attributed the high degree of perceptual synthesis found in these experiments to the highly predictive context from which the missing phoneme was taken. He proposed that phonemic restorations were the natural function of speech processing mechanisms which, in the normal course of events, had to deal with broken or masked auditory input. This conclusion is supported by Cherry and Wiley (1970) who used a similar paradigm to Miller and Licklider (1950), alternating speech sounds and silence. But, in their study, Cherry and Wiley used meaningful discourse whereas Miller and Licklider used word lists as stimuli. Cherry and Wiley found that eliminating everything but strongly voiced components of speech produced staccato sequences of very low intelligibility. This corresponded to the result achieved by Miller and Licklider in the "silent" condition. However, when noise of appropriate frequency was inserted into the gaps, the meaningful sequences became much more intelligible. Wiley (1968) confirmed this finding by removing the strongly voiced components and leaving the low energy speech sounds. He also found that the insertion of noise increased intelligibility of the meaningful sequences greatly.

Warren and Obusek (1971) interpreted these "multiple" phonemic restorations in meaningful sequences as strong support for the idea that the effect itself is the product of speech processing mechanisms which make extensive use of the syntactic and semantic context of the missing segment. To this end, Warren and Sherman (1974) performed an experiment to demonstrate the effect of subsequent context on this aspect of perceptual synthesis. In their study, information necessary for the correct identification of the missing phoneme occurred after the critical section containing the missing phoneme. Warren and Sherman also took the precaution of having the missing phoneme "mispronounced" on the original recordings. This was to ensure that no acoustic cues which could identify the missing segment were still present in adjacent phonemes. A typical sentence used by Warren and Sherman was "George waited for the deli(b/v)ery of his new colour TV". Here the deleted portion is shown between the brackets with the first phoneme as the original recording which was then deleted and the second as the phoneme which was reported by the subjects. Warren proposed that this restoration process was a "highly specialised form of auditory induction in which linguistic rules enter into the synthesis of the restored sound".

A much more striking example of this effect was produced by Sherman (1971) in an unpublished master's thesis. In his experiment, Sherman used extremely ambiguous stimuli which could only be disambiguated by subsequent context. Sherman systematically varied the subsequent context and observed the corresponding change in the subject's perception of the missing phoneme. The following is an example of this particularly elegant paradigm. The sentence "the *eel came off the ...." was ended by a number of different words - shoe, car, orange and table. (* Indicates the missing phoneme which was replaced with a broad band noise). As the subsequent context of the sentence varied, so subjects perceived different phonemes at the beginning of the stimulus "*eel". The percepts "h"eel, "wh"eel, "p"eel and "m"eal were heard, respectively, creating a semanticlly meaningful sentence.

This experiment is particularly interesting because it demonstrates that the mechanisms which restore missing phonemes in speech make extensive use of subsequent as well as preceding context, indicating that speech perception is not a strictly linear activity.

Up to this point, this synthetic aspect of the perception of missing segments in speech had been theorised as a product of high level processing mechanisms. Or, to put it another way, the phenomenon was conceived of as a concept-driven rather than a data-driven effect.

Evidence seemed to converge on this explanation with Sherman (1971), and Warren and Sherman (1976) emphasising the role of sentential context on the percept and the findings of Obusek and Warren (1973), which combined phonemic restorations with the action of another supposedly high level effect, that of verbal transformations (see Part 1 for a fuller discussion of this study). These findings fitted very well with the results of Warren and Obusek (1971) who found that the type of noise (e.g. cough, pure tone or white noise) which replaced the missing segment did not affect the level of perceived restoration. Thus, phonemic restoration was seen as an effect which could be influenced by high level processing variables but was much less prone to lower level data-driven manipulations.

This interpretation was eventually challenged. Layton (1975) found that all replacement sounds did not induce equal levels of restoration among his subjects. Layton discovered that white noise was superior to a pure tone in producing a restored percept. This finding cast doubt on the notion that only top-down features could seriously influence the reinstatement of missing phonemes.

The relative importance of top-down versus bottom-up variables was most incisively explored by Samuel (1981). Samuel adopted a new methodology aimed at disentangling the role of different variables by using signal detection theory as a framework within which to investigate phonemic restoration. In each experiment he produced two types of stimuli. In one, a phoneme had been 'replaced' by a noise, and in the other, the noise was added to a 'complete' word. The subject was asked to judge whether the stimulus was 'intact' or not. This allowed Samuel to calculate the subject's ability to 'discriminate' between the different types of item. The more this measure of discriminability (d') tended to zero, the less subjects were able to distinguish between replaced and intact segments, thus offering evidence for restoration. This method also produces a measure for response bias (Beta) which is the tendency of listeners to report either that all stimuli are intact or that none are intact. Samuel claimed that the failure of Warren and Obusek (1971) to find evidence to support any major influence of bottom-up variables in speech restoration arose because the replacement noises were always set at amplitude levels which would completely mask any phoneme even if it were present. Thus, the results represented a ceiling level of performance for any bottom-up influences which might be operating.

Samuel reduced the levels of replacement/added noise in relation to stimulus amplitude, thereby creating conditions under which signal detection parameters could be measured. He theorised that phonemes which acoustically resembled the replacement sound would give rise to lower d' scores using this method. Results confirmed this proposition, showing that the greater the similarity between the missing/intact phoneme and the replacement/added noise, the lower the subject's discrimination performance. Also, segments from real words demonstrated more restoration than those from pseudo words, and real words which had been primed performed best of all. This led Samuel to propose that "restoration (and more generally speech perception) depends on the bottom-up confirmation of expectations generated at higher levels". Warren (1982) concurred with this analysis. Although there are no detailed models of the way perceptual processes function when restoring missing phonemes, interactive approaches to this phenomenon, such as that of Samuel, are regarded as the most reliable interpretation of the available evidence.

Continuity in Non-Speech Stimuli

Miller and Licklider (1950) seem to have been the first to report the phenomenon known as "auditory induction". Here, a tone is repeatedly interrupted by a louder noise and subjects perceive the softer tone to be continuous. This alternation between softer and louder tones/noise has provided the classical methodology for research in this field. The loudness, duration and type of intervening noise in relation to the 'continuing' tone have provided the major variables under consideration in the majority of subsequent studies.

Although later work to some extent re-focused on the influence of 'context' (especially Bregman and Dannenbring (1977)), there has been a re-emergence of the original explanation of this phenomenon. That is, all of the early studies - Miller and Licklider (1950), Thurlow (1957) and Vicario (1960) - made use of Gestalt concepts in describing the nature of the experienced continuity, seeing it as analogous to various visual effects.

Miller and Licklider used the analogy of seeing the world through a picket fence; Thurlow used a static Gestalt image, likening the illusion to an auditory analogue of the figure- ground effect; Vicario called the continuity phenomenon the "acoustic tunnel effect", invoking a more dynamic Gestalt principle which maintains the existence of a moving object, such as a train, when direct viewing is temporarily obscured by a cutting or a tunnel. (It may be recalled that early explanations of the VT effect also focused on visual analogies). This type of explanation has been re-investigated much more recently by Bregman (1990) who also invoked Gestalt principles in placing this phenomenon in the much grander setting of a general theory of auditory perception.

It is more revealing to maintain a chronological perspective of research in this area in order to give a structure to the growth in ideas towards the present. In this respect, it was Thurlow who instituted a whole series of studies in this field : Thurlow (1957), Thurlow and Elfner (1959), Thurlow and Martin (1962), Elfner and Caskey (1965), Elfner and Homick (1966) (1967a) (1967b), Elfner (1969) (1971) and Thurlow and Erchal (1978). These studies manipulated a range of bottom up variables relating to the tone and interrupting sound, culminating in the functional conclusion which placed emphasis on, as Warren (1982) put it, "the facilitation by the louder sound of a continued firing of the neural units corresponding to the fainter sound .... arising through an excitatory post-synaptic potential". It can be seen here that, although earlier explanations made use of broad Gestalt principles to describe the illusion, more detailed explanations were built around bottom-up auditory theories.

One of the main features of the Thurlow and Erchal theory is that it did not require the louder sound to directly stimulate the units involved in the fainter sound, even though it was based on the notion of post-sound offset neural persistence. That is, it did not place any emphasis on the importance to the level of perceived continuity of an auditory match between the louder and fainter sounds.

This position was directly contradicted by Houtgast (1972), (1973), (1976), (1974b), (1974c), who maintained that the nature of the intervening sound is vital to the illusion. However, Houtgast's explanations were also in the bottom-up genre of lateral suppression on the basilar membrane. It was Warren, Obusek and Ackroff (1972) who, in agreeing with Houtgast about the nature of the louder sound, effectively broadened the base of how the illusion could be considered. They concurred that "the louder sound should be a potential masker of the fainter" but also proposed that the context of the sound was an important factor in the restoration process. To this effect, they proposed the following rule: "If there is contextual evidence that a sound may be present at a given time and if the peripheral units stimulated by a louder sound include those which would be stimulated by the anticipated fainter sound then the fainter sound may be heard as present". They demonstrated this proposition by alternating a softer with a louder sound on a 300 millisecond cycle which represented a long duration of interrupting noise in this sort of paradigm. The louder sound remained at a constant 1000Hz and 80dB while the frequency of the softer signal was varied between trials. Subjects were instructed to increase the amplitude of the fainter tone until it was perceived as 'discontinuous'. This essentially, is the point at which the continuity illusion disappears. Fainter tones at very low frequency levels (e.g. 150Hz) and very high levels were not perceived as continuous but, as the frequency levels approached that of the louder sound, subjects could adjust the amplitude up to 3dB of the 'masker' and still hear the softer tone as continuous. This experiment concluded that the nature of the louder sound is very important in establishing continuity. The closer the frequency match between the 'faint' and the 'loud' sounds, the stronger the continuity illusion.

This position concurs with that of Houtgast in finding that illusory continuity is enhanced when the louder sound is a "potential masker" of the softer sound or, as Bregman and Dannenbring (1977) put it, restoration is strengthened "in cases where the neural response to the louder sound would be most likely to interact with that of the softer sound".

Although the results of Warren, Obusek and Ackroff (1972) agreed with a bottom-up theory, their interpretation (confirmed by Warren (1970) with speech) differed from a neural response model based on 'continuous' tone by saying that 'any' sound might be restored, given sufficient auditory masking criteria, because the process was one of synthesis at a cognitive level. However, Warren (1982) drew a distinction between the processes involved in different types of continuity; 'Sounds' involved a "simple form of perceptual synthesis" whilst restorations of speech were "a highly skilled and complex type of synthesis drawing upon linguistic rules".

However, it was Dannenbring (1976) and Bregman and Dannenbring (1977) who pursued the notion that restoration in continuous tones was a synthetic activity rather than any variation of neural persistence. Although Warren et al had broadened the scope for theoretical re-interpretation of the illusion, they had not performed any experiments which could not be explained in purely bottom-up terms. In order to perform just such a critical study, Dannenbring (1976) used tonal glides. The underlying rationale was that in order to restore the missing sections of these auditory patterns, subjects would need to synthesise segments of sound which were either ascending or descending in terms of frequency. Persistent firing of the relevant neurons at the frequency offset would not account for a perception of continuity, as the onset frequency would be different. Subjects were asked to adjust the duration of the white noise until the sound pattern became discontinuous. In the longest glide used (two seconds) average noise durations were between 400 milliseconds and 500 milliseconds. These are very long periods of noise when compared to those used in other experiments in this area. Dannenbring observed high levels of illusory continuity using this tonal glide paradigm. This finding effectively refuted explanations of the phenomenon which are predicated on persistent neural activity, as firing which continued in cells dedicated to the offset frequency would not match up with the different onset frequency.

However, the ease with which subjects perceived the missing segments led to a follow up study by Bregman and Dannenbring (1977). Tones were 'ramped' in order to examine the effect of 'edge' information on the illusion. When these changes were introduced, the illusion of continuity was affected by the direction of the amplitude ramp. This demonstrated that subjects were sensitive to information derived from the positions just before noise onset and just after offset, the edges of the tone. For example, if this information indicated a reduction in stimulus amplitude prior to noise onset, then perceived continuity was lowered. Under these circumstances, the perceptual system seems to extrapolate the fall in amplitude into the interrupting noise, thus forming the hypothesis that the tone is less likely to be continuous. Having previously demonstrated that simple neural persistence would not be the sole cause of illusory continuity (Dannenbring (1976)), this experiment was an attempt to draw parallels with visual perception. In vision, edge information is extremely important when one object is perceived as continuing behind another. In demonstrating the salience of this information in auditory restoration, Bregman and Dannenbring claimed a clear analogy between the two effects to the same extent that the early researchers had used visual Gestalt phenomena to explain their results.

The culmination of this line of thinking rests with Bregman (1990) who has produced the most comprehensive integration of this phenomenon into a general auditory theory. Bregman has identified four rules which underpin the perceptual decision which is made on the continuity of a softer tone when it is interrupted by a louder noise:-

1. There should be no auditory evidence that the softer tone stops then starts again at the onset and the offset of the louder noise.

2. During the louder noise "some of the neural activity in the auditory system" is required to be "indistinguishable" from that which would have occurred if the fainter sound had really continued.

3. The auditory evidence available should strongly point to the conclusion that the two distinct presentations of the softer tone should come from the same source. According to Bregman, this would then place them in the same auditory 'stream'.

4. The transition of the softer sound in to the louder should be abrupt rather than gradual, thus mimicking the function of an interruption rather than one sound steadily changing in to another.

In order to support these rules, Bregman has undertaken an insightful survey of evidence which sets a description of this phenomenon in a larger context. It is beyond the scope of this thesis to make any sort of detailed critique of Bregman's rules - that would be more appropriate to a work dedicated to auditory aspects of the non-verbal continuity illusion. However, a few pertinent points should be made at this juncture.

Bregman proposes that the phenomenon of illusory continuity consists of two components, one deciding "whether" the signal continues or stops and the other answering the question as to "what" exactly has continued. Bregman contends that his four rules are sufficient to answer the "whether" component as this can be explained in terms of "primitive processes" which also function in primary stream segregation. He claims that the "what" part of the phenomenon, that is, the content of the illusory percept, is beyond the scope of his theoretical intentions. Bregman is not retreating to a new model of neural persistence but highlighting a belief in a theory of auditory perception which gives one set of processes the primitive function of assigning sounds to their sources, while higher schema-based processes apply an analysis from "learned knowledge".

The Continuity Illusion in Music

McAdams and Bregman (1978) in "Hearing Musical Streams" informally reported a phenomenon they called "gated music", which appears to be another term for illusory continuity. However, their brief observational note on the subject does not constitute any serious investigation. Prior to the experiments described here, there was only one report in the literature on the perceptual restoration of missing musical notes. That was the report of Sasaki (1980).

Sasaki (1980) looked at the phenomena of restoration and click migration together, in both speech and music. However, while some statistical evidence was presented on the localisation of the extraneous noise, no such data was provided for the continuity illusion. The results simply state, in one sentence, that "according to listener introspection" the missing notes had been restored. From these findings Sasaki concluded that music was similar to speech but qualitatively different from other forms of auditory induction. However, it is not made clear on what contingent base this proposition stands, as no other variables were manipulated in respect of either category.

There is currently one other paper in existence which deals with restoration in music, DeWitt and Samuel (1990). The publication of this study post-dated the experiments presented here. Hence, a discussion of the DeWitt and Samuel findings will be reserved until after the present experiments have been considered.

A New Framework for the Continuity Illusion

The aim of this section is to examine and clarify some of the basic elements of the continuity illusion from the perspective that the phenomenon is governed by the same processes in speech and all non-speech stimuli. It is not the intention here to construct a detailed model by reference to previous research. The object of this exercise is to confront afresh some of the basic questions about the nature of this creative mechanism and theorise about attributes which it must possess in order to function. It is in the light of this discussion that the experimental evidence presented later in this section should be evaluated.

The first issue to address, one which is all too often ignored, is why should any form of illusory continuity exist at all? The answer to this most basic question must lie in the illusion's ecological utility. The main use of such a mechanism must be to preserve the 'sense' of an interrupted auditory percept. Hence, there does not seem to be any reason to suppose that this should be less true of an everyday sound, which might for example foreshadow imminent danger for the listener, than for a section of verbal discourse. In both cases, the preservation of the sense of a percept through creative continuity during brief periods of interruption would have distinct advantages.

Placing the maintenance of meaning at the centre of a framework for the continuity illusion implies that bottom-up data-driven variables are of no importance. This is not the intention. The relevance of bottom-up factors is rightly highlighted by Bregman (1990). However, to concentrate on data-driven aspects is akin to looking down the wrong end of a telescope. A better perspective can be gained by examining Figure 4.1 which represents the general paradigm in all continuity studies:

Figure 4.1

T ------------------- →

[A______N / ✠____ B]

P

A = The auditory events preceding N.

B = The auditory events following N.

N = Replacement noise/silence.

✠= Missing/restored element.

T = Time.

P = Whole percept.

For listeners to experience restoration of ✠, certain bottom-up criteria have to be met. For instance, if N is represented by silence, then the sensory evidence would contradict continuity and the phenomenon will not occur (Miller and Licklider (1950)). Hence, data-driven factors have to be congruent with the possibility of restoration for it to take place. In this sense, the purely auditory facets of the stimulus are critical to the restored percept. However, these circumstances are essentially descriptive of 'when' continuity might take place but do not explain 'how' or 'why'. The role of bottom-up input should properly be regarded as a baseline for the phenomenon. All the correct auditory elements must be in place for the illusion to occur, but they do not represent a complete explanation; they are hence necessary, but not sufficient.

Given that auditory conditions are congruent with continuity and that the basic purpose of the system under consideration is to preserve the sense of the stimulus, the next question which arises is, what happens at the offset of A ? (Figure 4.1). Clearly, at the instant at which A stops and N starts, the perceptual system has no knowledge of what is to occur at B. But, logically, B must be known in order that the sense of the stimuli can be interpolated accurately. For example, let A = 'Trans'. This syllable can be completed and thus turned into a meaningful word in a number of ways:

Figure 4.2

A N B

Trans/✠/.......

/a/ ct

/i/ ent

/p/ ort

/f/ er

/l/ ate

In Figure 4.2, ✠ could be one of a large range of speech sounds which could only be correctly restored by the disambiguating segment B. This reasoning holds perfectly well for other complex everyday sounds, including music, and the conclusion which follows is that the process of restoration cannot be an entirely linear activity. Knowledge of the whole token is necessary for continuity to occur.

The concept of non-linearity in the continuity illusion is not new, especially in relation to speech (Sherman 1971). So, should it be asserted that essentially nothing happens at the offset of A until B disambiguates the percept? Not necessarily. In fact, it is much more likely that the system is active rather than passive at this point. The theory proposed here is that when confronted with an ambiguous situation, (the offset of A and the onset of N) the perceptual system proceeds to prime the range of possible alternatives for ✠ . This partial activation would serve the purpose of enhancing the speed at which the final percept is constructed once the nature of B is established.

The framework within which illusory continuity exists is the result of a set of processes which operate in real time, priming and preparing the system, on the basis of prior and current information, for probable incoming data, together with a set of processes which perform the retrospective function of summing up the available evidence (A + N + B) to create what is finally heard by the listener in perceptual time. The common purpose of the experiments in this section is to provide some evidence by which this framework can be evaluated.

CHAPTER 5 Experimental Series 2

Experiment 2.i

Introduction

The framework for the continuity illusion, which has been outlined above, highlights the maintenance of sense across an auditory percept. This general concept could apply to any interrupted token. However, musical stimuli seem to provide a particularly fertile place from which to begin to analyse the effect of top-down factors on perceived restoration. Musical sound provides a range of complex interrelations between individual events which evolve through time into a coherent whole. Hence, if this process is disrupted by an extraneous noise, and sense is to be preserved, it would follow that the level of restoration experienced by the listener could be significantly affected by a number of higher level cognitive factors, such as the familiarity of a theme or its metrical structure.

Therefore, the aim of this experiment is to examine the potential potency of different top down variables as critical factors in the restoration process. This approach is an attempt to shift the theoretical viewpoint from which non-speech stimuli have traditionally been analysed.

Method

Design

The two factors under consideration in this study were the familiarity of the musical theme and the metrical position occupied by the missing note. These variables are clearly related to cognitive structures beyond the physical representation of the token. That is, a theme can still be perceived as familiar if it is played, for example, at a high or low volume, and the metre of a piece can be discerned equally well from a flute or a piano. These factors rely on some aspect of prior knowledge in order to function. Hence, in examining the effect which familiarity and metrical position have on perceived continuity, this experiment is testing the notion that concept-driven variables can critically influence restoration.

This gave rise to the following design:

Familiarity;	Familiar		Unfamiliar	
Metrical Position;	Strong	Weak	Strong	Weak

One major problem with continuity studies has been how to measure restoration. Traditionally, subjects have been asked to respond `Yes' or `No' to the proposition that a softer stimulus continued during a louder interrupting noise. However, Samuel (1981) pointed out that this methodology did not take into account the possibility of response bias. That is, subjects could have a tendency to respond positively whilst not actually perceiving any continuity at all, or vice versa. In order to assess the possible effect of bias, Samuel re-organised the task into a signal detection paradigm where the primary stimulus really did continue during the noise in 50% of the trials. In a series of elegant experiments using this methodology, he was able to isolate the effects of response bias in restoration. However, the price of achievement of this laudable goal was the restrictions placed on listening conditions. These restrictions involved a reduction in the amplitude of the extraneous noise being either added to a complete primary stimulus or replacing a missing segment. The amplitude of the noise was reduced to a level at which subjects, when given the task of listening to just the critical segments (without the rest of the primary stimulus), could discriminate easily between the added/replaced conditions. That is, the level of the added noise had to be substantially lower than that of the main signal. Discrimination levels for this 'critical segment' task could then be compared with those obtained for the added/replaced conditions using the whole primary stimulus. This comparison gives the measure of restoration under this paradigm.

Although this method has produced some important insights into the continuity illusion, there is a logical paradox raised by the nature of the listening conditions. This paradox is that if these conditions were the only ones to exist in nature, then there would be no reason for a cognitive restoration mechanism to have developed at all. That is, if interrupting noises never reached an amplitude that would occlude a primary stimulus, then, by the principle of Occam's razor, less complex systems would be competent to deal with what would be a very simple auditory problem. Therefore, the present research was specifically and deliberately aimed at studying restoration under listening conditions which genuinely obscure the primary stimulus.

During piloting work, which involved extensive debriefing, subjects reported that their experience of continuity was often ambiguous and not accurately reflected in a`Yes'/ `No' response, and that the enforced use of a dichotomous response tended to introduce a consistent bias in responding. Hence, it was reasoned that a more flexible form of response would remove a central factor causing bias. Under the new paradigm, subjects could express their level of certainty of stimulus continuity on a rating scale from 1 - "Certainly did not continue" - to 7 - "Certainly did continue". This regime enabled a subject who was unsure about the exact nature of her/his percept to choose a response which more accurately reflected her/his subjective experience. This outlet for uncertainty gave the subject who was sure of stimulus continuity in most cases, but less convinced in others, the means of conveying the texture of the illusory percept, thus avoiding the over-simplification and bias that a series of `Yes' responses would yield. This would also be the case for a subject who made regular `No' responses when unsure. The use of this more flexible paradigm avoids this form of bias.

The experimental strategy outlined above has the drawback of not producing a measure for bias, as would a signal detection paradigm. However, the method proposed here does have three advantages : rating scales allow subjects to make a response which is more sensitive to the nature of their individual percept. This, in turn, leads to a reduction of bias caused by a `Yes'/`No' protocol. It also allows more freedom in selection and amplitude of the interrupting noise.

A further refinement adopted by this study was to introduce two distractor tasks in order to ensure that subjects spread their attention over the whole stimulus token, as a listener would under natural conditions, rather than concentrating on one task-oriented segment. These distractor tasks required the subjects to state whether a stimulus contained two consecutive identical notes, and to give an estimate of the duration of the whole token.

Stimulus Selection, Preparation and Presentation

Familiar themes were selected from a group of classical motifs which were piloted prior to the experiment. It was not a difficult process to assemble a number of themes that were familiar to all of the pilot group, as television and radio have incorporated classical motifs into their output, and, thereby, into our everyday lives, under a variety of guises. The unfamiliar themes were assembled by selecting an obscure motif from the work of each composer represented in the familiar group.

None of the pilot sample took part in the actual experiment. This precaution was taken to ensure that no subject had any exposure to the unfamiliar motifs. The danger with this protocol was that some of the real subject participants might not recognise some of the familiar themes or might know some of the unfamiliar themes. In order to avoid this eventuality, the debriefing process checked the correct categorisation of the motifs and no subjects had to be discarded for this reason.

Stimulus preparation involved entering each theme into a Steinberg Pro 24 MIDI Sequencing software Version 1.1 music programme via a Casio C Z 1 keyboard. The sequences were then edited by assigning a single note of each melody to a different MIDI channel so as to produce a broad band white noise of greater amplitude than the replaced musical tone.

The stimuli were subsequently presented to subjects via a cassette recorder and speakers, with testing taking place in group sessions. This format was chosen in preference to individual presentation via headphones in order to provide more realistic listening conditions. Listening with a group of other people forces a subject to be prepared for a wide range of auditory possibilities, which could be ignored if headphones were the only possible source of sound stimulation.

Subjects

Subjects were all undergraduates at the City University, London. Their ages ranged from 18 years to 32 years with an average of 22 years. The group of 28 consisted of male and female subjects. All subjects were paid for their participation. Subjects were enlisted on the basis that they had not undergone any specialist musical training beyond the Associated Board of the Royal Schools of Music, ABRSM, Grade 5. This stipulation was designed to exclude music specialists of all kinds. However, some musical training was allowed in order that the sample should reflect the level of knowledge in the population, and hence be as representative as possible.

Procedure

Subjects were tested in groups of not more than 10. The instructions informed subjects that they would be hearing a number of musical themes and they would have three tasks to perform after each musical presentation. Subjects were told to divide their attention equally between the tasks. Two practice trials were then given, after which subjects could ask questions. Every effort was made to ensure that all the subjects understood the instructions. After each experimental trial, subjects made their responses on an answer sheet provided, indicating their experience of continuity on a rating scale, and giving their responses to the two distractor tasks. For an example of these sheets see Appendix 2.

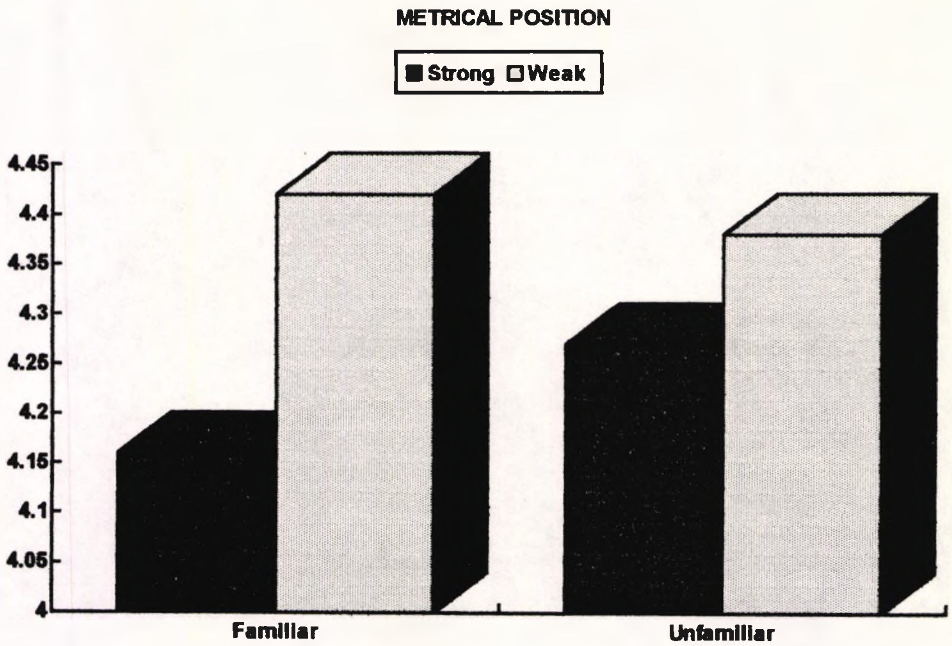
Results and Analysis

The results shown here are derived from rating scales on which subjects indicated how sure they were that a stimulus had continued during an extraneous noise (see Appendix 2).The mean rating scores for each stimulus category are given in Table 5.1 and graphically in Figure 5.1.

Table 5.1

		Familiarity	
		Familiar	**Unfamiliar**
Metrical position :	**Strong**	4.16	4.27
	Weak	4.42	4.38

Figure 5.1



The data were analysed by a 2-way, within-subjects ANOVA. Neither of the variables, nor the interaction between them, proved to be statistically significant. The initial factor to be considered, Familiarity, had no significant effect.

$F(1,27)=0.121 \quad p>0.05$

This indicates that there was no overall significant difference in the restoration ratings given to familiar and unfamiliar themes.

Metrical position also yielded an insignificant F value :

$F(1,27)=0.547 \quad p>0.05$

This shows that the metrical position of the missing note did not have a significant bearing on the level of continuity experienced by the subjects.

There was also no interaction between the variables of Familiarity and Metrical position:

$$F(1,27)=1.089 \quad p>0.05$$

While this study appears to have produced rather negative results in terms of the effects of the two variables on restoration, the results of rating scores suggest that subjects can experience high levels of continuity in interrupted musical stimuli.

Discussion

On a broad level of analysis, the results offer no support at all to the notion that familiarity or metrical structure have an effect on the perceptual process of restoration. However, to accept this conclusion would be to neglect the evidence which a closer inspection of the data provides. It is clear from Table A shown in Appendix 2 that there is some inter-stimulus variability. That is, some musical themes received much higher continuity ratings than others. This can be seen to most striking effect in unfamiliar theme 1, where a missing note from a strong metrical position only achieved the total rating of 52 from a possible 196, whilst a note from a weaker metrical position from the same theme was rated at 169/196(see Appendix 2, Table A). However, this strong / weak division was not consistent and an example which showed the opposite tendency was seen in unfamiliar theme 7 - strong 134/196, weak 66/196. Although degree of variability in the data is insufficient to undermine the assumption of homogeneity of variance underlying the analysis, there appear to be differences in the way individual stimuli perform as elicitors of the continuity illusion, and it is worth giving some further consideration in the following paragraphs to the way in which the variables may exert their influence.

It is possible that the two variables under consideration may not behave in a simple and unidirectional way in the restoration process. Take the example of 'Familiarity'. On one hand, it is possible to suggest that this variable should enhance continuity, as the perceptual system should "know what to insert" when the noise appears in order to complete the percept.

Conversely, it could be argued that prior knowledge of the theme may only serve to make the system more critical of a presentation which does not conform to the existing schema.

These two propositions should not be viewed as mutually exclusive. In fact, it is the contention here that both are true. That is, familiarity does have an effect on restoration but its action is complex and is highly interactive with other salient variables. There is some of evidence to support this proposal from the distribution of continuity ratings. For familiar themes, there are only three stimuli which achieve ratings of 75% and over and only one which falls below 50% (that being 49.5%), whilst six unfamiliar themes have ratings of 75% and over and five which are below 50%, the lowest of these being 26.5%. So, although the mean ratings for the two conditions are almost indistinguishable, the pattern of response in each case does seem to be different, with familiar themes less likely to attract very high or very low ratings whilst the unfamiliar themes tend to obtain more extreme ratings. It therefore seems that familiarity does have an influence on the nature of the percept, though this is masked when mean ratings averaged over tunes are used as an index.

Similarly, it can be asserted that musical structure (here in the limited sense that not every note is of equal value to the percept) is having an effect, as demonstrated by the fact that there are large differences between the restoration of two different notes from the same theme. However, this occurrence is not reflected with any consistency along the chosen dimension and it may be that the influence of structure has also been obscured by complicated interactions with other variables.

Overall, these considerations suggest that although the variables of familiarity and metrical structure may well play a role in restoration, the way in which they exert an influence is complex. Other variables may well be at work and these possibilities need to be teased out. This line of thought formed the basis of the next study in this series.

Experiment 2.ii

Introduction

Although the first experiment in this series failed to show a significant effect of the variables of familiarity and metrical position on restoration, the large inter-stimulus variability suggests that potential effects may have been masked by other factors. It is possible, for example, that other higher order variables were at work in the construction of the restored percepts. In fact, it is likely that restoration, particularly in real life listening situations, is the product of a number of interrelated and interacting factors which combine to create the emergent percept. In order to isolate the operation of any of these factors, it seems regrettable, but necessary, to reduce the richness of 'real' music and create stimuli specifically tailored to the investigation of a particular dimension. This is the approach adopted in this experiment.

The variables selected for investigation with more controlled stimuli were : predictability of the melodic line, and the metrical position of the missing note. Familiarity, in any real, long-term sense, could not be investigated since the stimuli would be specifically generated for the experiment and hence unfamiliar. Predictability was considered to be a factor worthy of further investigation on the basis of examination of the data of Experiment 2.i. On a close inspection of the themes which elicited the highest restoration ratings, a high proportion of them seemed to exhibit the common characteristic of having a smooth and regular melodic line. This characteristic, which can be broadly conceptualised as the cognitive feature 'predictability', was therefore selected for further study.

A second variable prompted by the data from Experiment 2.i was the relative importance of the missing note in the metrical structure of the theme. It was evident that different notes from the same theme showed clearly different levels of restoration. Hence, the position of the missing note was examined as a potentially significant influence on perceived continuity.

Method

Design

A two by two within-subjects factorial design was employed. The concept of predictability was defined here for operational purposes by constraining the degree of pitch variation between notes in any given melodic line. For example, a predictable melodic line is characterised in this context by small and regular variations in pitch between consecutive notes. While unpredictable melodic lines display abrupt disparities in the pitch of consecutive notes. Two levels of melodic predictability - High and Low - were examined, as were two levels of metrical position of the missing note - Strong and Weak.

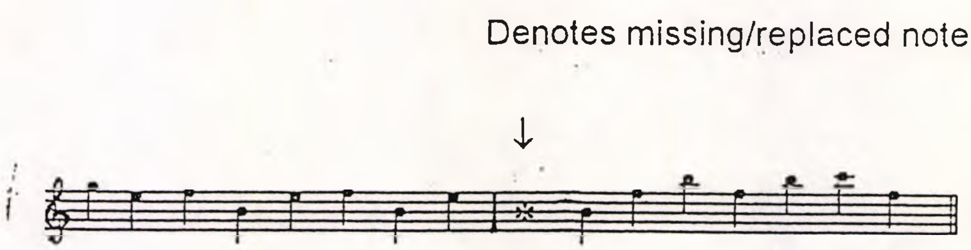
The protocol by which restoration was measured, that is, rating scales, was the same as Experiment 2.i. The same distractor tasks were also employed.

Stimulus Construction, Preparation and Presentation

Musical stimuli were constructed for this experiment using a set of criteria which were chosen in order to reduce the sources of uncontrolled variability. Predictable stimuli were constructed so that there was a small difference in pitch between one note and the next, while in the less predictable themes pitch differences between notes were randomly generated within the range of one octave. Examples of High and Low predictability stimuli are shown below.

Figure 5.2 High Predictability

Figure 5.3 Low Predictabiity



Subjects

Subjects were all undergraduates at the City University, London. Their ages ranged from 18 years to 28 years with an average of 22 years. The group of 35 consisted of male and female subjects. All subjects were paid for their participation. As in the preceding experiment, subjects were enlisted on the basis that they had not undergone any specialist musical training beyond ABRSM Grade 5. This stipulation was designed to exclude music specialists. However, some musical training was allowed in order that the sample should reflect the general level of musical knowledge in the population.

Procedure

Subjects were tested in groups of not more than 10. The instructions informed subjects that they would be hearing a number of musical themes and they would have three tasks to perform after each musical presentation. Subjects were told to divide their attention equally between the tasks. Two practice trials were then given, after which subjects could ask questions. Every effort was made to ensure that all the subjects understood the instructions. After each experimental trial, subjects made their responses on an answer sheet. (See Appendix 2).

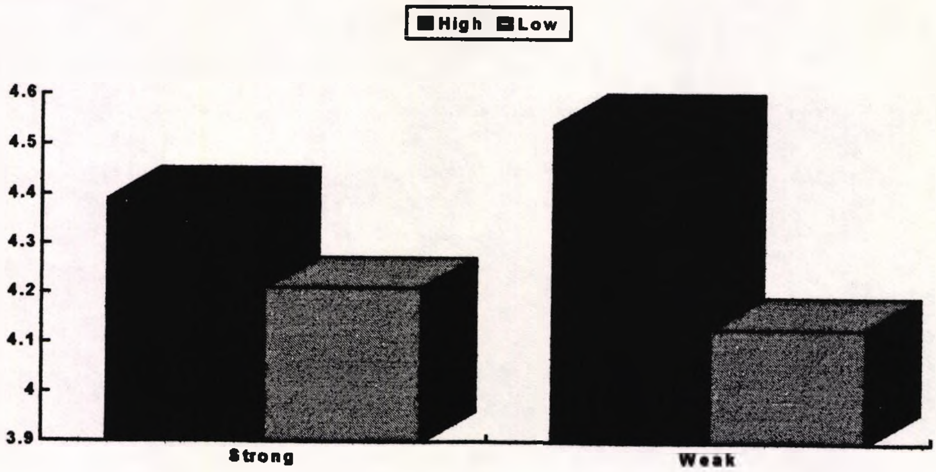
Results and Analysis

The results shown here are derived from rating scales on which subjects indicated how confident they were that the stimulus had continued during the extraneous noise (see Appendix 2). Table 5.2 and Figure 5.4 show the mean ratings obtained from subjects for each stimulus category in the design matrix.

Table 5.2

		Metrical Position	
		Strong	**Weak**
Melodic predictability	**High**	4.39	4.54
	Low	4.21	4.31

Figure 5.4



An ANOVA was used to investigate the statistical significance of the data. This revealed no significant difference in the continuity ratings given to missing notes from different metrical positions:

$F(1,34)=0.061 \quad p>0.05$

However, the analysis did reveal a significant difference between the two levels of melodic predictability:

$F(1,34)=6.240 \quad p<0.05$

This means that stimuli which followed a highly predictable melodic course elicited higher continuity ratings than the unpredictable tokens.
The interaction between the two variables was not significant :

$F(1,34)=0.662 \quad p>0.05$

Hence the effect of predictability is not dependent on a particular metrical position of the missing note.

Discussion

The results here indicated significantly higher ratings for stimuli which exhibit a highly predictable melodic line. This would appear consistent with the general hypothesis underlying this series of experiments, namely, that more conceptual factors are active in restorative processes. It could be argued that this discrepancy in the continuity experienced by subjects in the high predictability condition might not have been due to higher level factors or schemas, but could have been the result of a more primitive trajectory tracking mechanism, since the stimuli under consideration followed very simple progressions. This argument does have some intuitive appeal. However, in an exhaustive review of literature on this topic, Bregman (1990) finds that the balance of evidence cannot even confirm the existence of such mechanisms.

The 'trajectory' hypothesis is also undermined by a closer examination of the experiment itself. The replaced note in the metrically strong group of melodically predictable stimuli was also the first note of the new bar beginning the transposed repeat of the sequence. Hence, there was a discontinuity in trajectory at this point and no primitive tracking mechanism could have been of any use in restoring the missing segment. Only a retrospective overview of the whole stimulus, revealing the essentially repetitive structure, could have been of use at this point. These considerations lead to the conclusion that some higher level factors are at work in producing restoration. Some broader schema of predictability needs to be invoked as an explanation, rather than any notion of moment-to-moment trajectory tracking.

The need for a broader explanation is underlined by the observation that while there were differences between tunes of high and low predictability, tunes of low predictability, as indicated by the rating data, were nevertheless perceptually restored. This means that, although melodic predictability did make a significant difference to the level of perceived continuity, it should be seen as only one strand of a complex story as restoration was also clearly elicited by less predictable tokens.

In seeking a broader and richer explanation of this data it should be remembered that although there is a measure for the subject's confidence in the perception of continuity, there is no measure for exactly what is being heard along with the extraneous noise. In the melodically predictable condition, this would seem to be less of a theoretical difficulty, as the stimuli follow a simple pattern and the number of notes which would fit, in order to resolve the percept into continuity, is very small. It does not matter that one subject's perceptual system will complete the percept in a slightly different way from that of another subject, it only matters that the problem requires a simple melodic schema for good continuation.

However, there does seem to be a problem with the unpredictable stimuli. Here, there is no melodic structure at all which will provide a framework for restoration, and so the question of what the subject perceives impinges on the central theme of how the subject perceives a continuity. If there is no melodic information to predict the missing auditory event, then what influences the restorative process in this category? It could be that although there is no melodic predictability, each note in these tunes is of equal length and the temporal distance between them is also uniform. Hence, if each theme is taken as a pattern of auditory events, then, despite the variability of pitch, there is a form of regularity which must elicit the restoration of "an event" to make it complete.

The idea that there are many interwoven strands of processing which proceed in parallel is central to the explanation of perceptual continuity. This section of the thesis focuses on showing that top-down factors are vital in the perception of this phenomenon in non-speech stimuli, but it is the general contention that higher and lower order processes work simultaneously to produce all auditory percepts. This view will be expressed more fully in the general discussion at the end of this chapter. Thus, in relation to the present experiment it seems that the higher-order factor of predictability is important, but it is probably not the only factor to play a part.

Experiment 2.iii

Introduction

The aim of all the experiments in this section was to demonstrate different aspects of a non-verbal analogue to the phonemic restoration effect. The previous studies have examined various factors such as familiarity and melodic predictability on restoration. These factors were features of the stimulus which may or may not have been utilised by higher level cognitive processes in maintaining continuity. Rather than focusing on the various facets of the stimulus which might influence restoration, this experiment concentrates on a different aspect of the analogue. The approach was to formulate a hypothesis which was based directly on a functional framework of the process of percept reconstruction. Elements of such a framework were discussed earlier in this section. One of those elements was the requirement that any system capable of producing a continuous percept from disrupted input should be able to utilise segments of disambiguating stimulus information which occur after the interrupting noise. This aspect of the restorative process provided the focus for this experiment.

In speech research it has been shown that information which occurs after the noise can significantly affect the perception of a missing segment (Sherman (1971), Warren and Sherman (1974)). To some extent it could be argued that Dannenbring (1976) and Ciocca and Bregman (1987) have led the way in extending this to non-verbal stimuli. Dannenbring's aim was not to examine the effect of a specific piece of post-noise information on restoration, but to examine continuity when the frequency of a stimulus tone at the onset of an interruption is different than at its offset, as in the case of an ascending or descending tonal 'glide'. Conversely, Ciocca and Bregman examined the effect of specific segments of post-noise information by varying the 'exit' frequency gradients of steady tones from interruption offset. However, as Ciocca and Bregman's declared aim was to examine the ability of the system to track frequency trajectories, the effect of the post-noise sound did not purport to offer any analogue for the use of a specific disambiguating 'chunk' of information which is the basis of the speech based phenomenon.

The hypothesis underlying the present experiment is that the perceptual system is able to make use of specific information, occurring after the missing segment, in a functionally similar way for both non-verbal and speech stimuli, in order to adjust and restore the total percept. The aim of the present study is therefore to test this notion by providing disambiguating information about the missing segment, immediately or soon after the temporal position of the missing sound. The most informative level of this would be to repeat the segment containing the missing sound completely, with the missing sound present in the appropriate position, so that the total context of the missing sound is available. If the perceptual system can make use of this post-hoc disambiguating information, levels of restoration should rise relative to a situation where such information is not available.

In terms of relating this to speech restoration, it should be clearly understood that this paradigm did not intend to provide a mirror image of the Sherman or Warren and Sherman studies, as the notion of 'semantic reference' is absent. The aim was not to try to demonstrate that the non-speech and speech sounds are indistinguishable as stimuli, because that is plainly not the case, but that the perceptual system operates in a qualitatively similar way across the whole range of auditory events. So, although it cannot be shown that musical or other sounds are disambiguated on precisely the same semantic basis as speech, it may be possible to demonstrate that information, occurring after the critical segment, is utilised in a functionally similar way to change and clarify a clouded percept.

Method

Design

The experiment employed a two by two by three factorial design, with repeated measures on all factors.

The variables were:

1. Provision of post-noise disambiguating information, by repetition of the complete context of the missing note. In this experiment the context is defined as the two notes on either side of the missing note Two levels of this variable were used. Disambiguating information, in the form of an exact repetition of the original context segment complete with the missing note in the appropriate position, was either provided, or not. In the latter case, a set of different notes that did not repeat the structure of the original segment and so could not serve to provide disambiguating information about the original structure, were used in the comparable section of the musical line. Reference to Figure 5.5 will clarify this procedure.

Figure 5.5

2. The nature of the context from which the critical note was removed. There were two levels of this variable : a constrained context which was highly predictive of the missing note (Figure 5.6), and a less restricted sequence structure which did not prescribe the missing note (Figure 5.7).

Figure 5.6
Context predicting the missing note.

Figure 5.7
Non-predictive context

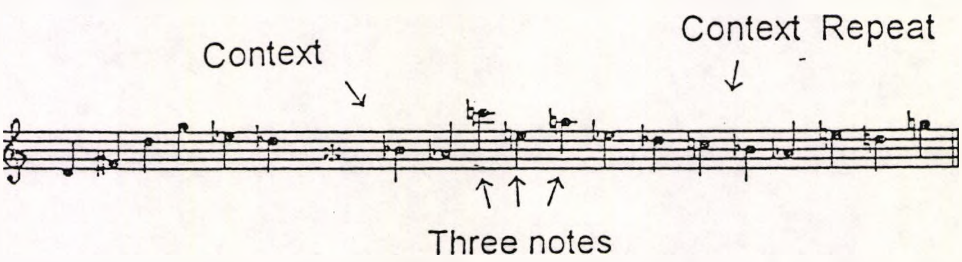
The purpose of this variable was to assess the processing system's ability to recognise different levels of disambiguating information. For example, the system may only be able to utilise post-noise context information which is highly predictive of the missing note.

3. Distance of the potentially disambiguating section of the stimulus line from the final note of the initial segment containing the missing note (Figure 5.8). There were three levels of this variable. The potentially disambiguating section either

- immediately followed the original with no other intervening notes, or
- three notes separated it from the original, or
- five notes separated them.

Figure 5.8

This example shows a three note separation interval.



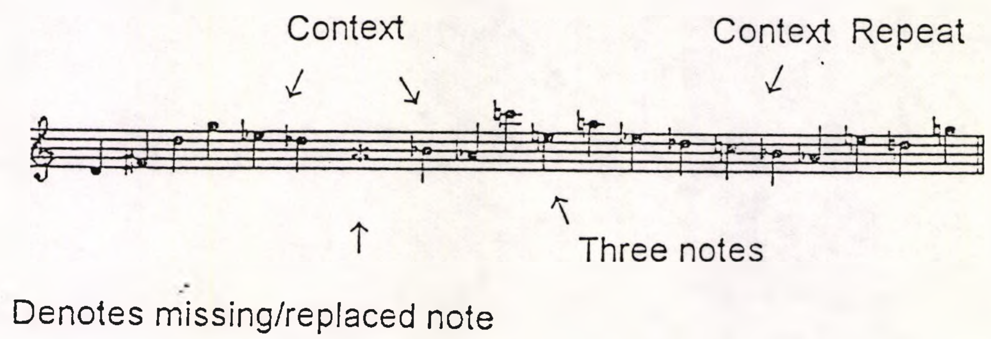
This variable examined the possibility that there is a limit to the distance at which disambiguating information can be utilised by the restorative processes in order to complete a disrupted percept.

In the case where no disambiguating information was to be provided (the second level of Variable 1) this element of 'distance' is nevertheless included in order to fulfil the practical constraints of the design.

Stimulus Construction, Preparation and Presentation.

Stimuli were constructed using an Atari ST 1040 computer, Steinberg Pro 24 Version 1.1 software and a Casio C Z 1 keyboard. The synthesised sound used in the experiment approximated to that of an acoustic piano. Stimuli were constructed in order to comply with the criteria of the experimental design.

Figure 5.9



A further point which should be noted here is that although the stimuli consisted of notes which were, in their frequency spectra, generated to sound like those of a musical instrument (piano) and the ordering of the notes conformed to some extent to musical forms, the final themes (because of experimental constraints - see Design) did not sound very musical. This is of no real consequence to the theoretical thrust of this thesis, but may have some implications for anyone who wishes to draw any specifically "musical" conclusions from the results, and with respect to this, some care should be exercised.

Subjects

Subjects were all undergraduates at the City University, London. Their ages ranged from 19 years to 27 years with an average of 22 years. The group of 33 consisted of male and female subjects. All subjects were paid for their participation. Subjects were enlisted on the basis that they had not undergone any specialist musical training beyond ABRSM Grade 5. This stipulation was designed to exclude music specialists of all kinds. However, some musical training was allowed in order that the sample should reflect the general level of musical knowledge in the population.

Procedure

The experimental procedure was the same as that in the previous studies presented in this section. Subjects were tested in groups of not more than 10. The instructions informed subjects that they would be hearing a number of musical themes and they would have three tasks to perform after each musical presentation. Subjects were told to divide their attention equally between the tasks. Two practice trials were then given, after which subjects could ask questions. Every effort was made to ensure that all of the subjects understood the instructions. After each experimental trial, subjects made their responses on an answer sheet. (See Appendix 2).

Results and Analysis

The results shown here are derived from rating scales on which subjects indicated how sure they were that a stimulus had continued during an extraneous noise (see Appendix 2). and the mean continuity ratings obtained from subjects for each stimulus category, on which the analysis was based, are shown in Table 5.3.

Table 5.3

Disambiguating information :		Presented		Not presented	
Type of context :		Predictive	Non-predictive	Predictive	Non-predictive
Distance of	0 notes	4.6515	4.8030	4.3636	3.6364
disambiguating	3 notes	4.3788	3.9394	4.0152	3.6667
information	5 notes	4.6364	3.2727	3.6364	3.4394

Tables 5.4 and 5.5, and Figures 5.10 and 5.11, show data derived from Table 5.3, but split into two sections for clarity of presentation and discussion. Table 5.4 shows only variables 2 and 3 - the predictiveness of the context surrounding the missing note, and the distance of the potentially disambiguating section from the original - at only one level of the first variable - the provision of disambiguating information. The data are restricted to cases where disambiguating information **was** provided; that is, where the initial segment was repeated exactly, with the missing note in position. To consider **distances** of a potentially disambiguating segment

from the initial one, when the former does **not** carry disambiguating information, makes little sense.

Table 5.4

Distance	**High predictive context**	**Low predictive context**
0 notes	4.652	4.803
3 notes	4.379	3.939
5 notes	4.636	3.273

Figure 5.10

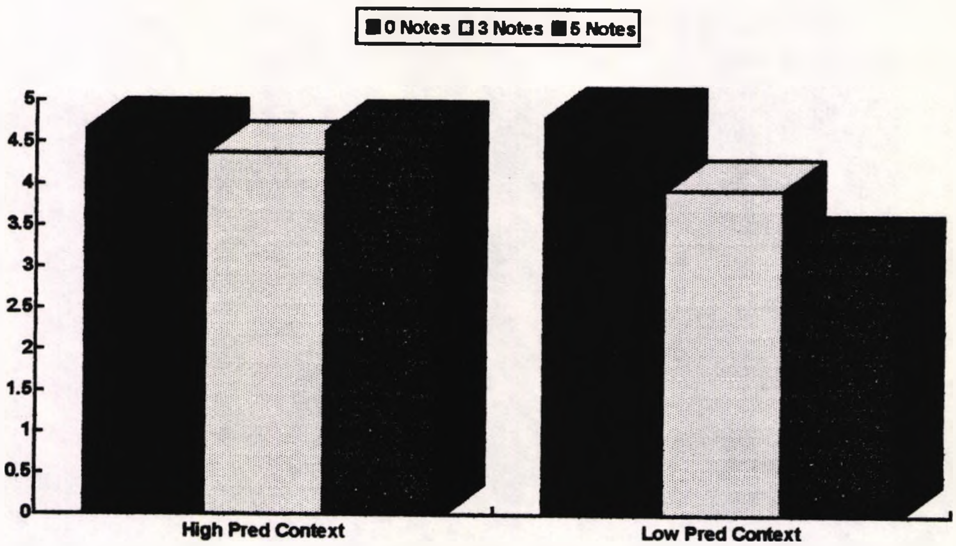
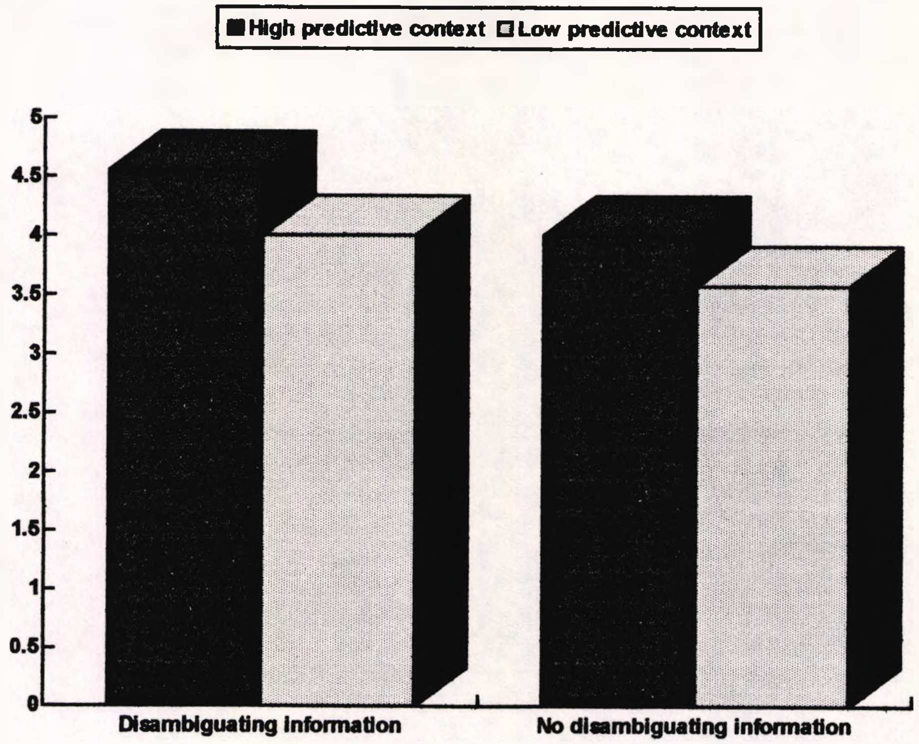


Table 5.5 and Figure 5.11 show the effect of type of context and presentation of disambiguating information, collapsed over different distances of the potentially disambiguating segment from the original.

Table 5.5

	Disambiguating information	**No disambiguating information**
High predictive context	4.556	4.005
Low predictive context	4.005	3.581

Figure 5.11



The results were analysed by a three-way, repeated measures ANOVA. The primary object of this study was to examine the effect on restoration of repeating the context of the missing note at some point after its original presentation, compared with restoration in the absence of post-hoc disambiguating information. This is embodied in Variable 1 - 'the provision of post-noise disambiguating information' (See Design section), and yielded a significant F value:

$$F(1,32)=4.838 \quad p<0.05$$

This indicates that continuity ratings are significantly higher for those stimuli which contain a repetition of the context of the missing note. This means that subjects are able to make use of specific segments of disambiguating information which occur after the interruption to enhance the perception of continuity.

Although the effect of context repetition is clear, it has a rather complex relationship with the other variables. This interaction will be disentangled after the other main effects have been reported.

The second variable under consideration was predictiveness of the context from which the missing note was removed. There were two levels of this variable - High and Low predictiveness, and the manipulation produced a significant effect:

$$F(1,32)=12.726 \quad p<0.01$$

This indicates that restoration of a missing note is facilitated by a predictive context, and hence that part of the restorative process is based on an appreciation of the immediate auditory environment.

The third variable was the distance from the initial presentation at which disambiguating information could be provided. There were three levels of this variable - zero, three and five notes between the end of the original presentation and the start of the potentially disambiguating segment. The effect of this variable proved to be highly significant :

$F(2,32)=5.465 \quad p<0.01$

The general implication is that the greater the distance between the two segments, the weaker the restoration.

While none of the two-way interactions was significant, the analysis revealed a significant three-way interaction between the variables:

$F(2,64)=4.802 \quad p<0.05$

The reason for this interaction can be seen by reference to Figure 5.10 and Table 5.3 (above) . Figure 5.10 shows a two-way interaction between Variable 2 and 3 - 'predictiveness of context', and 'distance of disambiguating segment' - at one level of Variable 1 - that of provision of disambiguating information. It is clear from Table 5.3 that, at the other level of Variable 1- where no disambiguating information is provided - no similar interaction between Variables 2 and 3 is indicated.

The interaction revealed in Table 5.3 is supported statistically by a sub-set of critical pairwise comparisons within the main analysis. There is no significant difference between restoration levels for predictive and non-predictive contexts when the disambiguating information (repetition of segment) is provided immediately (0 note distance).

$F(1,32)=0.308 \quad p>0.05$

At a distance of 3 notes the difference between the two types of context is significant at the 0.05 level :

$F(1,32)= 5.588 \quad p<0.05$

At a distance of 5 notes the difference between the contexts is highly significant :

$F(1,32)=18.725 \quad p=0.01$

The interaction clearly shows that where the context of the missing note is highly predictive, then the disambiguating information remains equally useful whether it occurs 0, 3 or 5 notes from the offset of the original presentation. However, when the context of the missing note provided a less predictive setting, then the distance at which the sequence was repeated gained importance. If the information was immediately available, restoration was improved, but, as the intervening interval increased, the effect of provision of disambiguating information decreased.

Discussion

The results of this study lead to some fairly straightforward conclusions and some which require a little more interpretative effort. Of the clear results, the most important is that stimuli in which the note context is repeated receive significantly higher continuity ratings than those in which it is not heard again. This is an unequivocal indication that the perceptual system is able to make use of specific segments of information which occur after the missing segment to enhance the imperfect percept. This is the most important aspect of this experiment and it means that the non-speech analogue of restoration can be asserted with even more confidence. However, the other findings of this experiment speak less directly to the notion of an analogue, offering a more complex commentary on the functions of the phenomenon.

The second variable under consideration was the predictive nature of the note context from which the missing note was removed. The results indicate that this was a highly significant manipulation. Notes removed from a very simple linear sequence, that is, from a predictable note context, were much more likely to be restored than notes which were removed from less consistent environs. This is very much in line with the result obtained in Experiment 2.ii, in which the more predictable themes produced higher levels of restoration than did the unpredictable stimuli. However, it should still be noted, that regardless of type of predictive context, the stimuli with the post-hoc repetition received higher continuity ratings than those in which no disambiguation was provided. This suggests that the variables which contribute to restoration are cumulative in the way in which they contribute to the final percept.

However, the level of salience attributed to each variable seems to be mutable depending on the circumstances. This is borne out by the third variable, the distance of the repeated section from the original. In the predictive context condition, the distance of the repeated sequence from the original did not affect the perceived continuity, in that all of the stimulus categories benefited roughly equally. That is, it did not matter if the repetition of the context occurred immediately or three or five notes later; the system was still able to make use of the information. In the condition of low context predictability, only immediate repetition had a significant effect on restoration, raising the level of perceived continuity. However, this effect was so strong that the restoration levels demonstrated by this subset of stimuli were significantly higher than those demonstrated by the stimuli with a predictive note context but no repetition.

The picture which emerges seems to be that a predictive note context has a positive and significant effect on continuity which can be raised even further if subsequent disambiguating evidence is available. The effect of the predictive context in raising the likelihood that the percept will be restored means that information occurring even at some temporal distance from the missing note can be utilised. However, if the note context does not make a very restrictive contribution to the identity of the missing event, then only information in the immediate proximity of the disruption can be seized upon to disambiguate the percept. The implications of these findings will be discussed in a broader framework at the end of this section.

Experiment 2.iv.

Introduction

The aim of the preceding experiment was to examine the effect of specific pieces of post-hoc information on restoration. This tested one of the propositions about the nature of the restorative process asserted in earlier in this section. Another part of that framework was that information occurring before the interrupted section of the stimulus could also affect the perception of the missing segment. The theory was that relevant prior information would prime the system, stimulating related representations and making the missing element more available for restoration. Hence, the present study can be viewed in a similar way to Experiment 2.iii as an examination of a framework of ideas on the functioning of the processes which underlie the continuity illusion. Whereas Experiment 2.iii explored the effect of post-hoc information, this experiment focuses on information preceding the critical portion of the stimulus.

A secondary aim of this experiment was to broaden the base of the analogue for non-verbal sounds in restoration paradigms, beyond what might be described loosely as musical sounds, into auditory stimuli which may be encountered by subjects in everyday situations. This was an attempt to match experimental conditions with real life auditory experience where any type of sound can be interrupted or obscured for brief periods.

Method

Design

The experiment involved subjects listening to a series of non-verbal sounds, such as might be encountered in everyday life, from which segments had been deleted and replaced with bursts of noise.

A between-subjects design with three levels of a single variable was employed. The variable was the amount of information given to the subject prior to the listening task. There were three conditions to which subjects could be assigned:

Condition 1: Subjects were not given any information about the stimuli prior to the listening task.

Condition 2: Subjects were primed by hearing all the stimuli in a "complete" state (no missing segments) prior to the main listening task. The stimuli were never named by the experimenter - subjects only received auditory information.

Condition 3: Priming in this condition was elaborated by subjects listening to the complete sounds and, after each one, writing a brief description of what they heard. The experimenter then gave the name of each stimulus. This was considered to be a much stronger priming regime because the act of describing the sound would ensure that subjects concentrated on the internal structure and composition of each stimulus.

The extent to which subjects experienced restoration when listening to the sounds was measured on a rating scale, as in previous experiments.

Stimulus Construction, Preparation and Presentation.

The stimuli for this experiment were recorded with an Aiwa cassette recorder and then sampled by a Korg DSS 1 synthesiser - sampler in order to manipulate the section of sound to be removed and replaced with a broad band white noise. The stimuli were then recorded onto two separate cassette tapes, one with complete versions of the sounds, the other with a brief segment of each removed and replaced with a broad band white noise. The sounds were presented to subjects in small groups via a Sony TC - FX3 cassette recorder and speakers, with the complete stimuli being played prior to the test session as part of a priming regime in Conditions 2 and 3.

Subjects

Subjects were all undergraduates at the City University, London. Their ages ranged from 19 years to 31 years with an average of 23 years. The group of 39 consisted of male and female subjects. All subjects were paid for their participation. Subjects were enlisted on the basis that they had not undergone any specialist musical training beyond ABRSM Grade 5. This stipulation was designed to exclude music specialists of all kinds. However, some musical training was allowed in order that the sample should reflect the general level of musical knowledge in the population.

Procedure

Subjects were tested in groups of not more than 10, each group being randomly assigned to a priming regime. Subjects in Condition 1 received no information prior to the test session, while in Condition 2 the complete versions of the stimuli were presented without any listening task and, in Condition 3, subjects were asked to write a brief description (a couple of words), of each sound, which was subsequently named by the experimenter.

After each priming regime, all subjects received the same instructions for the test sessions in which the sounds with missing/replaced segments were presented. The instructions informed subjects that they would be hearing a number of sounds and they would have three tasks to perform after each I presentation. Subjects were told to divide their attention equally between the tasks. Two practice trials were then given, after which subjects could ask questions. Every effort was made to ensure that all of the subjects understood the instructions. After each experimental trial subjects made their responses on an answer sheet, see Appendix 2.

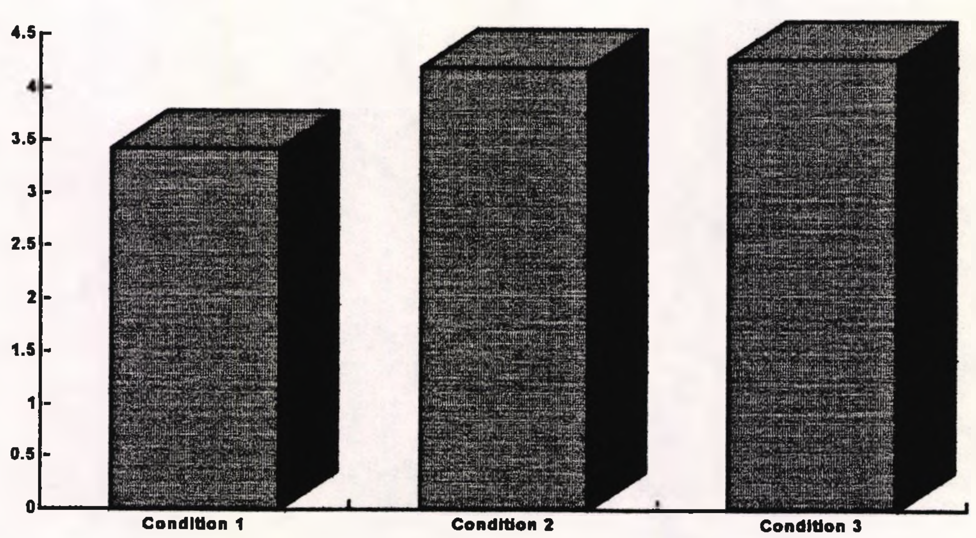
Results and Analysis.

The results shown here are derived from rating scales on which subjects indicated how sure they were that a stimulus had continued during an extraneous noise (see Appendix 2). Table 5.6 and Figure 5.12 show the mean ratings obtained from subjects for each category of priming regime.

Table 5.6

	Condition 1	Condition 2	Condition 3
Continuity rating	3.4250	4.1986	4.2846

Figure 5.12



An ANOVA was used to examine the effect of different levels of priming on the perceived continuity of the stimuli, and revealed a significant effect:

$$F(2,38)=4.644 \quad p<0.05$$

This indicates that the information which the subjects received before the test session did affect the degree to which they restored the missing segment of sound. However, if the data are examined more closely using pairwise comparisons of the means for each condition, a clearer picture of the effect of priming emerges:

Condition 1 vs Condition 2 $F(1,38)=6.359$ $p<0.05$

Condition 1 vs Condition 3 $F(1,38)=7.561$ $p<0.01$

These comparisons show that Conditions 2 & 3 differ significantly from Condition 1. This means that both of the conditions which provided a priming regime were more effective in eliciting restoration than the control condition. However, the two priming conditions were not different from each other. This indicates that an increase in the processing demands made by the regime in Condition 3 does not increase the level of restoration.

Discussion

The analysis of the results obtained in this experiment clearly supports the hypothesis that priming increases the chances that the perceptual system will accept an interrupted token as continuous. This finding fits in with the framework for the continuity illusion outlined earlier in this section. The framework suggests that relevant information which precedes an interrupted sound and provides stimulation for connected internal representations facilitates the perceptual system's ability to infer the content of the missing fragment of sound. This concept will be explored in greater detail, in the context of the other results which have been obtained in this series of experiments, in the General Discussion which follows.

It was not a surprise to find that the experimental conditions which used a priming regime elicted higher levels of restoration than the control condition. However, it was surprising to discover that the elaboration of the priming regime in Condition 3, which placed a more explicit burden on the subject in terms of related processing tasks, did not induce levels of restoration which were significantly higher than the simple exposure to the complete sounds which occurred in Condition 2. It was expected that an increase in activity in relevant connected representations would follow an increase in the processing tasks and, hence, an increase in the levels of perceived continuity. However, this did not prove to be the case. There are two possible reasons for this; the first is that extra elaboration of the priming regime simply has no effect on restoration. It may be the case that once the system has been alerted to the possibility of an impending sound, then further elaboration does not increase the potential to restore an interrupted stimulus. In these circumstances, the prime can be regarded as an all or nothing agency, 'switching on' the system as preparation for related input, but not raising that level once the system has been alerted. The second possibility is that, although Condition 2 did not provide any explicit elaboration in the priming regime, subjects participating in a laboratory experiment may have paid a greater degree of attention to the stimuli in this situation than if the same sounds had been heard in an everyday context. Hence, the subjects themselves could well have provided an increase in the elaboration of the processing allotted to the sounds.

There is no critical piece of evidence which can give an indication as to which of these explanations is correct. However, to accept the second reason would mean assuming a fairly uniform degree of unprompted processing elaboration among the subjects. This is not impossible, but this would also imply similar levels of attention and motivation within the group. Therefore, since the first hypothesis is simpler in that it does not require so many variables to be working together at the same time, it will be favoured here as an explanatory model of the results.

CHAPTER 6

General Discussion - Experimental Series 2

There were two main aims of the experiments in this section. The first was the affirmation of a non-speech analogue for the phonemic restoration effect. This was explored primarily by investigating the influence of concept-driven variables on perceived continuity in Experiments 2.i and 2.ii and by examining processing parallels in Experiment 2.iii. The second aim was to test the framework for the functioning of the continuity illusion proposed in this section. This was largely investigated in Experiments 2.iii and 2.iv. In order to examine the progress towards these aims and to arrive at a broader picture of the continuity illusion, this discussion will begin with a brief recapitulation of the findings of each experiment.

The initial findings of Experiment 2.i did not demonstrate clearly any effect on restoration of the concept-driven variables, familiarity or the metrical position of the missing note. However, on closer inspection of the data, patterns of variability emerged, giving an indication that high level variables were influential in non-speech restorative processes, but their action was difficult to evaluate as the complex musical themes used as stimuli gave a range of cues to the perception of continuity. In short, the rich nature of the stimuli did not allow the particular variables under consideration to be sufficiently isolated to yield an unambiguous effect. The conclusion then was not that concept-driven variables were unimportant in restoration, but that the experimental technique was insufficiently precise to disentangle a highly interactive process.

The solution to this problem was to use specially constructed stimuli. In Experiment 2.ii, stimuli were produced with the aim of reducing the number of concept-driven variables which could be available to the processing system. This approach demonstrated that melodic predictability significantly increased restoration, but the metrical position of the missing note still appeared to have no effect on perceived continuity. The data from this study clearly indicated that if the correct concept-driven variables are identified, then their effect on restoration can be measured. Given that the influence of concept driven variables can be isolated and examined, then clearly the continuity illusion in non-verbal stimuli is not solely mediated by auditory factors, so, at this level of analysis, an analogue for phonemic restoration can be confidently asserted.

Evidence obtained from Experiment 2.iii also converged on the concept of a non-verbal analogue. However, this experiment approached the proposition from a different angle, examining the nature of the process rather than the nature of the mediating variables. Experiments using speech have shown that the process of restoration involves the utilisation of information occurring after the missing segment. Experiment 2.iii was designed to test for this process in the resolution of disrupted non-verbal stimuli. This experiment discovered that post-hoc information is used by the system to resolve an ambiguous percept with non-verbal tokens, in much the same way as in speech. This finding offers further confirmation of the analogous nature of the continuity illusion in speech and non-verbal stimuli.

However, the aim of Experiment 2.iii was not just to establish a possible non-verbal analogue, but also to test the propositions advanced earlier in this Section and to look more closely at the way in which the processes involved in restoration identify and interpret post-hoc information. The main features of the rather complex data revealed that the processing system was strongly influenced by the context of the missing note. A highly predictive context produced better restoration and enabled the system to utilise that disambiguating information at some temporal distance from the interruption. In contrast, the contexts which were not highly predictive of the nature of the missing note were only utilised by the system if repeated immediately.

Experiment 2.iv also concentrated on the processes involved in restoration by examining the role of information occurring before the stimulus token. It was hypothesised that the greater the degree of priming, the greater the degree of internal elaboration and hence more restoration. However, the analysis of the data showed that although priming the stimulus tokens did significantly raise restoration levels, increasing the amount of priming did not cause a corresponding increase in perceived continuity.

So what picture of the continuity illusion emerges if the findings from these experiments are integrated? The evidence presented above indicates that in order to restore missing elements in non-verbal tokens, knowledge-based systems are in operation. In order to sort out what 'knowledge' is salient, it is important to comment on any variables which did not appear to aid continuity. In this respect, the lack of attention paid by the system to the metrical position of the missing note is worthy of comment. This variable failed to have any significant effect in Experiments 2.i or 2.ii. It could be argued that relatively untrained subjects simply do not perceive the metrical structure in music. However, evidence from Experiment 1.ii, where subjects reported significantly more illusory transformations of a stabilised musical theme when a note of greater metrical importance was replaced with a noise burst than with notes of lesser metrical value, and from Experiment 3.i, to be reported below, where extraneous clicks migrated to purely metrical boundaries, indicate that metrical structure is perceived by untrained subjects. Therefore, it can be assumed that although subjects are at some level processing metrical structure, it does not appear relevant to restoration. This is in contrast to melodic predictability, which does seem to influence restoration, indicating that the system places more value on information which is directly indicative of continuity of the whole, rather than emphasising the metrical value of one element. The stronger the evidence that there is an unfolding and relatively predictable pattern of events, then the greater the likelihood that the system will restore a missing or degraded element in order to preserve the 'meaning' of the whole stimulus.

However, there is evidence which might, on the face of it, directly contradict this interpretation. This evidence comes from the only other concerted study on this topic, DeWitt and Samuel (1990), which was published after the completion of the experiments presented in this thesis. Using a signal detection methodology, DeWitt and Samuel reported an insightful and incisive series of studies on restoration in music, producing data which appear to show that melodic predictability had an adverse effect on continuity - highly predictable tokens produced less restoration. This finding is in direct opposition to the results of Experiment 2.ii.

This raises the question of which study has the more valid results? The answer which will be asserted here is that both sets of findings are a true reflection of the restoration process. This proposition only appears odd if the variable of melodic predictability is viewed as unidirectional. Both studies indicate that the variable is salient to the processes engaged in restoration, one in a positive way, the other in a negative direction. This paradox can be explained if melodic predictability is seen, in relation to the central function of the restoration process (which was set out earlier in this section), to preserve the 'sense' of a degraded percept. Hence, the variable of melodic predictability may, under certain conditions, provide evidence to convince the system that continuity during an interruption is the optimum resolution of the percept, but it is possible to conceive of conditions which are so predictive of an auditory event that the system raises its threshold criteria for accepting disrupted data, making it more difficult to reinstate occluded input. The system is sensitive to information regarding melodic predictability and, up to a point, this information has a positive effect on restoration while it 'implies' the identity/existence of a missing fragment, but, when that information reaches a point at which it precisely constrains and defines the missing fragment, then the system is forced to be more discriminating towards the degraded input, with the result that a further increase in predictability diminishes restoration.

Evidence which supports this interpretation can be seen in a phonemic restoration study by Samuel (1981a) in which 'predictability' of non-words was increased by "cueing" - presenting the complete test item immediately prior to the onset of the interrupted version - which consequently increased restoration. However, when the 'predictability' of words was increased by using whole sentences as constraining contexts, restoration was decreased. This example demonstrates that a variable which affects restoration can produce contradictory results in perceived continuity. However, there is not really sufficient evidence in the non-verbal sphere to support or discredit the proposition. It can only be stated that this line of speculation offers the best fit for the limited data which has currently been reported, and replications and extensions of the research must be awaited before the situation can be further clarified.

One area where some agreement is apparent between DeWitt and Samuel and the experiments presented here is in the emphasis placed on restoration within a scale. Experiment 2.iii compares two 'contexts' for a missing tone, one of which was a foreshortened rising scale which proved to be significantly better at eliciting restoration than the non-scalar alternative. DeWitt and Samuel also demonstrated that information derived from scales produced an increase in restoration. However, they also found high restoration levels in segments removed from major chords. They attributed both findings to basic units of musical knowledge.

The validity of this aspect of DeWitt and Samuel's conclusions is beyond doubt. For instance, it was surely an apprehension of basic scalar continuity in Experiment 2.iii which enabled the system to restore the missing note. DeWitt and Samuel emphasised the implications for our understanding of the structure of music perception, and they also claimed that their findings supported the Fodor (1983) modular theory of perceptual architecture. The theoretical consequences of fully pursuing this line of argument are beyond the scope of this discussion, but they will be considered in more detail in the final section of this thesis, when all the experimental evidence has been presented. For the moment it is only necessary to note that the stem of the DeWitt and Samuel argument relates to the "low level" of the particular mediating variables - major chord and scalar context - which gave positive restoration data.

The description of this phenomenon will now move away from particular variables, which may or may not promote restoration, to the consideration of the elements which characterise the functional systems of restoration. This form of analysis was employed in Experiments 2.iii and 2.iv. To consider this, reference to Figure 6.1 may be of use:

Figure 6.1

T----------------->

[A______N/✠______B]

P

A = The auditory events preceding N.

B = The auditory events following N.

N = Replacement noise/silence.

✠ = Missing/restored element.

T = Time.

P = Whole percept.

Some basic predictions were made earlier in this section about the functioning of the systems which restore missing segments of a token. These predictions included the active state of the system at the offset of A and the non-linear processing of elements of B to account for information which could disambiguate ✠. Results from Experiment 2.iv demonstrated that priming which stimulated processing at the offset of A could increase restoration. These data were in line with the findings of Experiment 1.ii, which found higher numbers of illusory transformations clustered around a missing note which had been replaced with noise and used as a stabilised auditory stimulus. This seemed to indicate that the system was activating interrelated representations as part of a restorative strategy. Hence, it was postulated that greater elaboration during the priming stage would work on this strategy and improve restoration. However, while the general principle of priming increasing restoration proved successful, the idea that greater elaboration would further increase perceived continuity did not seem to be supported. It appears therefore that under normal listening conditions, the information derived from a prime acts as an 'on switch' in preparing the system for related input, and the effect observed in Experiment 1.ii is a magnification of normal processing strategy caused by the repetition inherent in that paradigm.

The other prediction made earlier was that specific fragments of information which occur after the offset of N can be utilised by the system to resolve uncertainty in the percept. Once again, the underlying principle was clearly supported by the evidence, but there were some unexpected aspects of the findings. Informally, it was suspected that the greater the distance of the critical context information from the missing note context, the greater the difficulty for the system in incorporating it into the percept. However, this was not the case for the highly predictive scalar note contexts. These improved restoration equally as well whether they were presented 0, 3 or 5 notes after the initially interrupted contexts. This can be contrasted with the contexts which were less predictive of the missing note, which only positively affected restoration when presented immediately after. As in the case of speech (Sherman (1971), Warren and Sherman (1974)), the system which produces restoration can function in a non-linear way but, for non-verbal stimuli at least, there is a case for maintaining that not all information is considered equally. It seems that some types of data concerning the missing fragment are more readily assimilated than others.

PART 3

Click Migration

Click migration refers to an auditory phenomenon which occurs when a listener mislocates an extraneous noise which has been superimposed on a spoken sentence. Extraneous noises, typically clicks, are said to perceptually 'migrate' to boundaries which are important in segmenting the stream of input.

In previous sections, the existing evidence which directly related to non-speech analogues of the phenomena under investigation was very slim indeed and so relatively more time was devoted to discussing the effects in speech studies. But when considering the phenomenon of click migration in musical stimuli, there is a small but significant body of existing work which should be addressed in some detail. The discussion of this auditory effect in speech will therefore be relatively brief and correspondingly greater consideration will be allotted to the studies which directly relate to the core of this thesis.

CHAPTER 7

Click Migration in Speech

The physical reality of connected speech is that it consists of a continuous and highly complex flow of auditory information. However, the listener perceives this information as a discontinuous series of words, clauses and sentences. Therefore, the auditory flow has to be recoded into representations which reflect linguistic structures. It follows that an essential feature of processing is the segmentation of the stream into meaningful units.

Click migration is a phenomenon which has been used experimentally to investigate the nature of segmentation in speech processing. The first example of this in the literature was Ladefoged and Broadbent (1960). This study demonstrated that subjects mislocated the superimposed clicks at places prior to their actual positions.

However, the most important of the early studies was Fodor and Bever (1965). Fodor and Bever theorised that the cause of the click's migration was the tendency of the perceptual system to maintain the integrity of discrete internal representational units. Fodor and Bever were the first to investigate the correspondence between possible linguistic parsing descriptions of a sentence and the perceptual segmentation of the stimulus. They discovered that a click embedded in a spoken stimulus sentence tended to 'migrate' towards the closest syntactically structural boundary and away from its actual position.

The nature of the perceptual elements proposed as processing units have for the most part been defined in purely structural terms, as constituents or clauses, on a surface versus deep structure level. Garrett, Bever and Fodor (1966) showed that brief noises tended to be localised at boundaries between the main clausal elements of a spoken sentence. Further evidence supported the notion that the speech stream is segmented around the major consituent boundary but the argument remained as to whether the perceptual units correspond to deep 'sentoids' (Bever, Lackner, and Kirk (1969) and Fodor, Bever and Garrett (1974)) or surface structures (Chaplin, Smith, and Abrahamson (1972)).

Carroll and Tanenhaus (1975) and Carroll and Bever (1976) challenged the clausal hypothesis, as it was known, by claiming that it was an incomplete description of segmentation. They claimed that it was not the structural clause but smaller functional clauses which were the basic segmentational unit. In order to formalise this view Tanenhaus and Carroll proposed a hierarchy of segmentation boundaries, with simple sentences at the top and adverbial subordinate and relative clauses towards the bottom. Further refinements were added to this position by Flores d'Arcais and Schreuder (1983). However, for the purposes of this thesis it is enough to understand that linguistic boundaries also form perceptual units to which extraneous clicks can be seen to migrate.

There is one other aspect of click migration research in speech which has some relevance in the context of this thesis, that is the tendency of extraneous noises to be localised at boundaries marked by prosodic features of the utterance. The vast bulk of research has concentrated on grammatical markers of speech segmentation. However, Wingfield and Klein (1971) demonstrated that patterns of intonation were important in processing streams of verbal information. The role of prosodic markers was supported by Wingfield (1975). Hence any non-verbal analogue for speech in terms of segmentation processes would require consideration of both grammatical and intonational factors.

Evidence for a Non-Verbal analogue for Click Migration

There is a small body of literature which relates to click migration in non-verbal stimuli, specifically in music. The first attempt at producing evidence for a non-verbal analogue appears to be Gregory (1978) who used six very short musical motifs as stimuli. The sets of notes were played in the same way in each condition the boundaries to which the clicks were supposed to migrate were entirely created by the visual representation of the music. That is, the notes were shown as either two groups of three or three groups of two. In order to create some form of perceptual set, subjects were encouraged to think of the music as it was represented. Hence, the boundaries under consideration were not inherently musical, but were created by visual representation and verbal instruction.

The results showed significant displacement of the 'clicks' towards the boundaries and it was also observed that these extraneous noises were perceived as occurring later than their actual position, whereas in speech they are generally perceived earlier. This experiment has received some criticism from Stoffer (1985) on the grounds of musical validity. Stoffer claims that the stimuli used were not long enough to reflect extended structure and the boundaries which were used were visual rather than auditory and thus did not cast any light on the notion of perceived musical segments. On the basis of a purely musical interest in the results, these comments are justified and, to be fair to Gregory, he recognises as much by suggesting the use of "longer passages of genuine music" in further experiments. But perhaps a more general conclusion can be drawn from this study in that it makes the first move towards affirming the reality and mutability of a perceptual 'chunk' outside the speech domain.

A more sophisticated attempt to demonstrate a musical analogue for click migration in speech was performed by Sloboda and Gregory (1980). This experiment used longer stimuli which were manipulated in order to reflect physical and structural boundaries. The results demonstrated that clicks were attracted towards musical boundaries which were indicated by both physical and structural markers, physical markers only or structural markers only, thus offering clear evidence for the psychological reality of musical segments. In presenting these findings, Sloboda and Gregory pose a question which is central to this thesis "...it seems pertinent to ask whether the systems developed in man for the comprehension and production of language are not subserved by non-specific cognitive mechanisms which can be deployed in a number of diverse ways including the perception and production of music?".

Although this experiment offers strong evidence for the similarity of segmentation of music to that of speech, there are two aspects of the study which prevent the analogue being complete. The first is acknowledged by Sloboda and Gregory who state "it is difficult to separate the effects due to auditory factors from those due to the inspection of visual transcripts". To some extent, this is a similar methodological problem to that of Gregory (1978), in that the interaction between visual cues to segmentation and auditory factors is uncontrolled. Furthermore, it is highly likely that the role of visual cues was compounded by the use of trained music students who would be familiar with interpreting musical scores in terms of phrase structure and emphasis. In short, although the study appears to support the notion that segmentation in music does occur, the difficulty of separating musical and auditory cues, especially in highly trained subjects, means that a general analogue between speech and music cannot be asserted beyond the restricted circumstances of this study.

Compared to Sloboda and Gregory (1978), Sasaki (1980) provides a much less incisive investigation into click migration. In fact, the study was primarily concerned with restoration. Sasaki was only interested in the "mis-location of clicks" rather than their migration to a segmental boundary. However, it does offer a chance to discuss one aspect of this phenomenon which has yet to be examined - are clicks perceived before or after their actual position? Sasaki, who directly compared speech and music stimuli, found that in both cases the click was perceived as prior to its actual position. This may be contrasted with Gregory (1978) who reported that his subjects heard the click later. This he ascribed to Titchener's (1909) Law of Prior Entry, where the aspect of the stimulus which is the primary focus of attention is processed first, that is, the melody is processed before the click.

However, Sloboda and Gregory (1980) clearly felt unhappy about the findings of their study which also found that clicks were perceived later. They theorised that, as the extraneous noise was positioned to coincide invariably with the centre of a note rather than its onset, this asynchrony caused the perceptual system consistently to assign the click to a later position, since more importance is given to the onset of a sound rather than its centre when making decisions about order. This led Sloboda and

Gregory to "re-calibrate" their results, adjusting scores by 50 milliseconds, in order to cope with the asynchrony, and this allowed them to pronounce that, given this adjustment, music, like speech, shows a tendency for the clicks to be perceived earlier than their actual position in time. This theory, explaining the late perception of clicks in music as a methodological problem (asynchronous onsets), may well be valid, but the post-hoc adjustment of results does not provide adequate evidence for it, and only, at best, some encouragement for further empirical testing.

The final study to be considered also fails to provide a firm answer to this question. Stoffer (1985) focused on the representation of phrase structure in musical stimuli and did not offer any conclusive opinion on an overall early/late perception of clicks in music or speech. Stoffer's concern in his rather elegant study was to establish the relative salience of different boundary points which were predicted within a given musical structure. The result was extremely clear, showing "correspondence between the formal description of phrase structure and the representational structure of the perceived music". That is, the distribution of click position judgements showed that the highest frequency of localisations occurred at the main phrase boundary, with lower order boundaries having significantly less attractive power.

Although click localisations show a pleasing adherence to the structure of the music, demonstrating segmentation in line with the theoretical parsing of the themes, the method by which Stoffer achieved this result should be borne in mind. In order to ensure that his subjects not only understood the notion of structure in music, but also the specific type of structure that they would encounter in the experiment, he gave students, who were already musically experienced, four lectures in the syntactic structure of German folk songs - the style of the music used in the experiment - and only those who passed the subsequent test with a score of 85% or above were chosen for the study. Whilst there is nothing wrong in principle with ensuring that subjects understand the material about which they will be asked to make judgements, this level of pre-test training should give pause for thought about the general applicability of the results to a wider population. Certainly, the results of the experiment show that those particular subjects segmented the auditory stimuli in ways which closely relate to the syntactic structure of the themes used. However, it is difficult to decipher whether this is a reflection of a normal way of parsing musical stimuli or a reflection of the specialised training they received.

Stoffer made some fairly forthright criticisms of Gregory (1978) and Sloboda and Gregory (1980) by saying that their experiments were flawed in allowing musically sophisticated subjects visual access to the scores of the stimuli under consideration, thus conflating the segmental effect across modalities. However, it could be argued that Stoffer's own extensive subject training programme conflates two ideas, that is, what segmenting rules it is possible to learn then apply, with more general rules which may be applied by the general population. This does not mean that the evidence should be disregarded. On the contrary, it can be seen as a strong example of how higher order information can feed into and modify the perception of auditory events. However, the degree of consonance achieved here between syntactic structure and click localisation may not obtain in more natural listening conditions.

The difficulty with using untrained subjects for this type of migration study is largely methodological. It lies in the problem that, without knowledge of musical scores, it is difficult for individuals to make a coherent response. In attempting to solve this problem, Stoffer performed a second experiment (reported in the same paper) which used reaction time. Strictly speaking, this method is not a click migration technique as subjects are asked to respond as quickly as they can when they hear an extraneous noise or click. Hence, it may more accurately be described as click detection. The advantage is that, as no visual array is ever seen, no knowledge of musical scores is necessary to make a response. However, it should be remembered that the theoretical base of the enquiry is also shifted; with click migration the theory is that extraneous noises are 'removed' to segmental boundaries arising during the creation of a perceptual unit, whilst in click detection the assumption which is made is that the system can respond more quickly to a noise at a boundary point than one which occurs between points. It is clear that both forms of investigation are directed at the same processes of segmentation, but it is not quite so clear whether both paradigms always operate under exactly the same rules. After all, responding as quickly as one can to a click seems to call on a slightly different set of resources than sorting out the exact position of that click amongst an array of other information. So perhaps results from this study should be considered as parallel to, rather than interchangeable with, migration studies.

Stoffer, however, considered the two types of segmental investigation as directly comparable. Yet, when results of his second experiment are compared to the first, it can be seen that there is a failure to replicate the degree of correspondence between syntactic musical structure and click localisation. The most interesting aspect of the second study was that the musically untrained listeners seemed to change their perception of musical boundaries through the experiment. Initially, the low-competence group seemed to segment the themes in a way which showed little relation to the musical structure, whilst the high-competence group demonstrated a segmental pattern which was to some extent correspondent with standard parsing. By the end of the experiment, the low-competence group was giving responses that were similar to the high-competence group. This seemed to indicate a type of informal learning through exposure to a great many musical stimuli of similar structure during the experimental procedure. It should be made clear that the musically unsophisticated group were 'segmenting' the themes which were presented to them, even in the early stages of the experiment, but not precisely in the way which the structure would predict.

In conclusion, there are some questions which need to be tackled if a clear non-speech analogue is to be claimed for click migration. The prime target is to examine migration in subjects whose musical experience can be seen as representative of a wider population. This must involve some solution to the problem of facilitating a clear response without contaminating the result with too many visual cues to segmentation. A secondary aim would also be to give a more decisive answer to the question of whether clicks are perceived as occurring early or late. Therefore, it will be the purpose of the rest of this section to pursue these points experimentally.

CHAPTER 8 Experimental Series 3

Experiment 3.i

Introduction

The purpose of this experiment was to pursue the notion of a non-speech analogue of the click migration phenomenon. Past studies have demonstrated that there is a strong likelihood that musical stimuli are segmented in a similar way to words. However, the strongest evidence which has been produced so far has relied on expert listeners with either a score to consult (Sloboda and Gregory, 1980) or who have been rigorously instructed in the structural syntax of the sorts of themes under consideration (Stoffer,1985). So, although there are some strong indications of click migration in non-verbal stimuli, there is inadequate evidence to claim an analogue. To establish the existence of an analogue less equivocally, it is necessary to demonstrate migration in subjects who are not musically trained. After all, click perception in speech is not limited to students of linguistics. The present experiment therefore focused on examining the segmentation behaviour during listening to music of subjects displaying a more representative cross-section of musical ability.

The methodological problems associated with this approach involved devising a practical way in which unsophisticated subjects could make a response without being biased by a visual display which in itself offers a segmental interpretation. The practical solution for these problems is discussed in the design section. Of greater relevance, in terms of a theoretical assertion of an analogue, was the problem of which elements of the stimuli could be used as segmentation markers and, in terms of this phenomenon, which elements would prove most and which least attractive to migrating clicks.

The approach which was used in this experiment was to look at the types of boundary markers which had proved most effective in speech research. This meant constructing stimulus conditions in which syntactic and prosodic features could be clearly studied. The aim was to separate out the different facets of the stimulus which may act as boundary markers and compare their relative salience. Also relevant was the point at which subjects were able to make a response - that is, at the end of, or during, the stimulus presentation. The purpose of this aspect of the study was to highlight any differences between the perceptual utility of different factors. For instance, could purely syntactic boundaries prove important before the whole structure is complete? It might be that the unsophisticated listener cannot interpret unemphasised grammatical segments without access to the complete theme.

Method

Design

The musical counterpart of grammar in speech was taken to be metrical structure of musical stimuli, and that of spoken intonation to be performance intonation. This contention is supported by Lerdahl and Jackendoff (1983), whose generative theory incorporates metrical structure as a central component in their construction of a grammar of tonal music. The individual and combined influence the variables, grammar and prosody, on click localisation was investigated. A structural boundary was provided by constructing a set of musical themes which divide metrically around a central point. This basic structural boundary could be emphasised by intonation, which was instantiated by variation of amplitude of critical notes.

The experiment employed two variables : type of boundary marker, with three different boundary definitions being used, and time of response indication - either immediate, at the moment of hearing the click (Response A), or retrospective, upon the completion of the stimulus (Response B). This gave rise to a three by two design, with boundary definition being manipulated as a within-subjects variable and time of response as a between-subjects variable.

The second variable, that of time of response indication, allowed for examination of the effect of whole or partial knowledge of the structure of the theme, since only in the retrospective response condition could the structure of the stimulus be fully realised before a response was required. The interrelation of this dimension was examined with the different types of boundary marker.

The three levels of the first variable - type of inter-segmental boundary marker - gave rise to three conditions:

Condition 1. A purely structural boundary was used. This was realised by a completely isochronic rendition of the stimulus - each note was of equal amplitude and duration. This eliminated all boundary features except the structural shape of the theme itself.

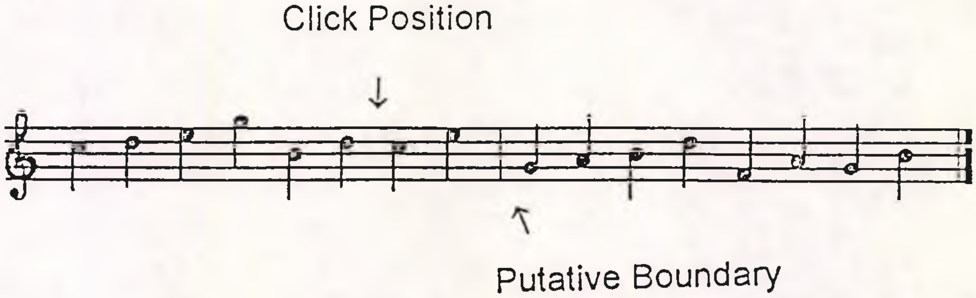
Condition. 2. In the second stimulus condition the boundary was defined by both surface and structural markers. The existence of a structural boundary was emphasised by an increase in the amplitude and duration of the first note of the sequence, and the first note of the second half of each theme. The aim of this was to mimic the normal interpretative emphasis given to the first note of a new bar. Hence, the physical cue and the structural boundary point were coincident.

Condition. 3. The boundary was defined by surface intonation alone. This involved emphasising a structurally inappropriate note. Since a structural boundary existed at the centre of the line, this intonational boundary was positioned to be as far removed as practical from the central point.

The aim of this condition was to examine how important purely physical features of a stimulus are in segmentation.

A set of six melodic stimuli composed of isochronic notes and conforming to the same basic structure - a metrical boundary dividing two melodically identical but pitch-different phrases - was generated. Each line was of approximately 5 seconds duration. A click of equal volume to the notes was superimposed in the line, in a pre- or post-boundary position, the click-to-boundary distance being constant at 4 units of input (a unit being defined as a note or inter-note interval).

Figure 8.1

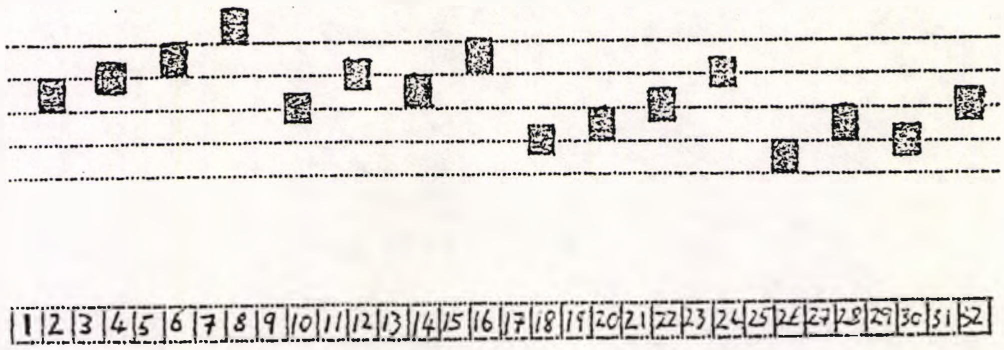


In order to ensure that subjects spread their attention evenly across the whole stimulus, two distractor tasks were provided. These tasks entailed asking subjects to rate the stimuli on two scales; Musical - Unmusical and Interesting - Boring. Distracting subjects in this way ensured that they could not concentrate their attention exclusively on the rather unnatural task of pin-pointing a click, but had to consider the whole auditory array.

Response Indication

Response indication was the major practical problem which needed to be overcome. Some method had to be found of allowing musically unsophisticated subjects listening to a brief musical stimulus to indicate at what point they had heard an extraneous click. After considerable time devoted to piloting various possibilities, the following paradigm was selected. Subjects were shown the musical stimulus in diagrammatic form (see Figure 8.2) with notes and inter-note spaces numbered consecutively, and were asked to indicate the number of the position at which they thought the click occurred.

Figure 8.2 - Diagram of Response Indicator



This representation shown in the above figure of the stimulus gives a clear indication of the melodic line, whilst the removal of all musical notation (except the stave lines) means that no visual cues are given as to how the stimulus might be segmented. As the representation appeared on an overhead slide, subjects could give the position at which they perceived the click by selecting a number, each digit corresponding to either a note or a space between notes. Hence, for example, should a subject feel that the click was coincident with the fifth note, then position 10 would be selected; if the point at which the extraneous noise was localised was between the seventh and eighth notes, then position 17 would be chosen.

The reason for displaying the representations of each theme on an overhead slide rather than simply using an individual response sheet, was that control could be exercised over the point at which the subject could make a response. Hence, in the response condition which allowed the subject to make a choice about the position of the click during the tune, the representation was displayed throughout the trial, whereas in the condition where the subject could only indicate his/her decision after the theme was completed, then the representation was only shown at the stimulus offset.

Stimulus Construction, Preparation and Presentation

The stimuli were created specially for this experiment, each using the same structural formula (Figure 8.1). The formula involved generating an initial theme which was then repeated at a different pitch. Stimulus preparation involved entering each theme into a Steinberg Pro 24 Version 1.1 music programme via a Casio CZ1 keyboard. The stimuli were then edited by entering a click sound which was always positioned at the same distance from a potential boundary marker.

Figure 8.3

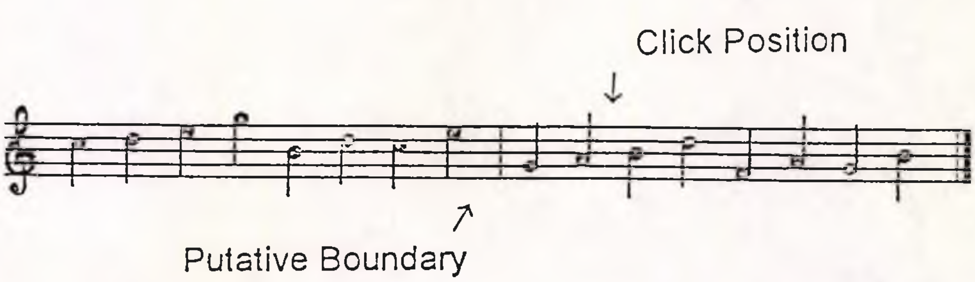


Figure 8.3. indicates the distance at which clicks were placed from the boundary. It was necessary for each theme to be played twice in each condition - once with a click placed before the putative segmental point and once with the click after. The reason for this was that if there is a tendency for clicks to be perceived early or late then this should apply equally to both click trials of any given theme. If no attraction is exerted by the theoretical boundary, then the distance between the click placements of the subjects should not vary significantly from 8. Therefore, when a comparison of the themes is made, if the difference between the judgements of the click positions is significantly less than 8, the effect of the segmental marker can be detected, despite any overall bias towards early or late perception.

A further point should be made about the positioning of the click. In all cases, this was placed between notes rather than occurring simultaneously with a note. This avoided the problem which Sloboda and Gregory (1980) theorised as the cause of late click perception in music. They claimed that clicks which occurred at the centre of a note were perceived late due to the note-click onset asynchrony. This potential source of confusion does not arise if the click never coincides with a note.

Subjects

Subjects were all undergraduates at the City University, London. Their ages ranged from 18 years to 26 years with an average of 22 years. The group of 16 consisted of male and female subjects. All subjects were paid for their participation. Subjects were enlisted on the basis that they had not undergone any specialist musical training beyond ABRSM Grade 5. This stipulation was designed to exclude music specialists of all kinds. However, some musical training was allowed in order that the sample should reflect the level of knowledge in the population.

Procedure

Subjects were tested in groups of not more than 10. The instructions informed subjects that they would be hearing a number of musical themes and they would have three tasks to perform after each musical presentation. Each theme was displayed as an overhead projection, (described above in Response Indication). In the immediate response condition the projection of each stimulus was visually available during the whole trial to enable subjects to respond at once when the click occurred. Conversely, in the retrospective response condition the overhead representation of the theme was not displayed until the stimulus offset, thus making it impossible for subjects to respond until the end of the trial. Subjects were asked to give a number corresponding to the position of the extraneous noise, and to give ratings for the subsidiary tasks. Subjects were told to divide their attention equally between the tasks.

Two practice trials were then given, after which subjects could ask questions. Every effort was made to ensure that all of the subjects understood the instructions. After each experimental trial, subjects made their responses on an answer sheet, shown in Appendix 3.

Results and Analysis

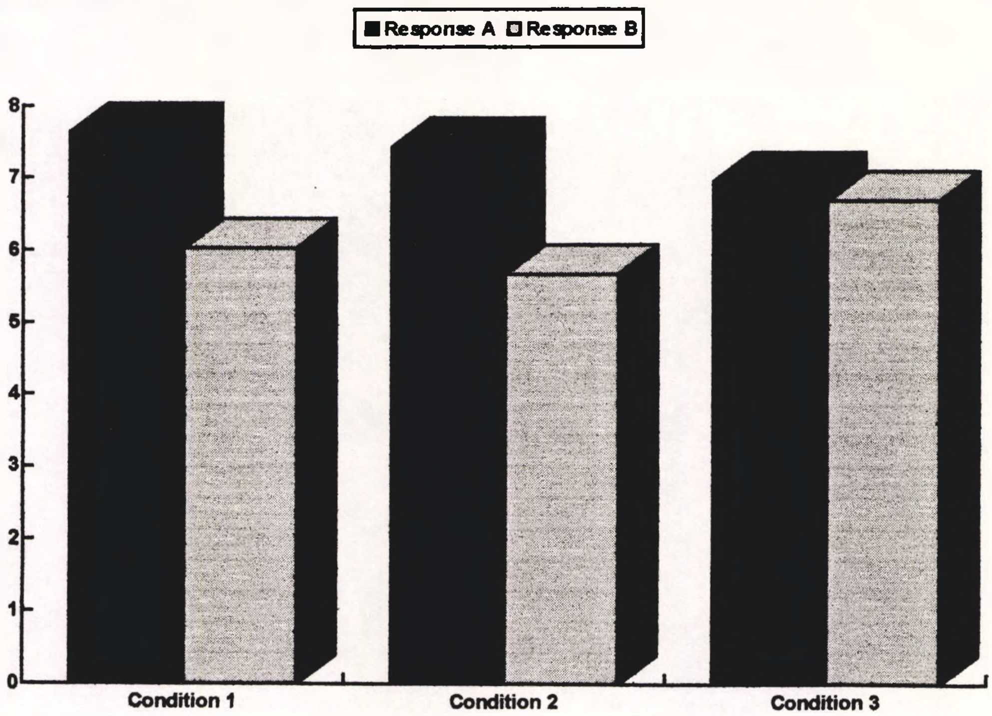
The results shown here are based on the concept that each stimulus was played twice, once with a click placed before and once after each boundary. The clicks were always placed four positions from the boundary. Hence, if the data from any stimulus is compared across the two presentations and if the positions of the clicks have been identified accurately, then the difference between the two click locations should always be 8 positions. If the putative boundary attracts the clicks then the difference should be significantly less than 8. This method of measurement avoids any difficulty which might occur in interpreting the data if there were any unquantified tendency to perceive the clicks earlier or later than their actual position. Table 8.1 shows the mean differences observed between the two click positions of the same stimulus, in each condition of boundary definition.

This is also graphically displayed in Figure 8.4.

Table 8.1

	Condition 1	**Condition 2**	**Condition 3**
Response A	7.63	7.42★	6.94★
Response B	6.03★	5.67★	6.68★

Figure 8.4



The symbol ★ indicates where the mean displacement differs significantly from the score of eight which would indicate accurate placement. This analysis was performed using one sample t tests. The values for t are listed below in Table 8.2. At 15 degrees of freedom $t > 1.753$ for $p < 0.05$ and for $p < 0.01$ then $t > 2.602$.

Details of t-values values and their associated probabilities are listed below :

Table 8.2

Condition 1	Response A	$t_{15} = 1.58$ $p > 0.05$
Condition 2	Response A	$t_{15} = 2.95$ $p < 0.01$
Condition 3	Response A	$t_{15} = 4.27$ $p < 0.01$
Condition 1	Response B	$t_{15} = 2.81$ $p < 0.01$
Condition 2	Response B	$t_{15} = 4.76$ $p < 0.01$
Condition 3	Response B	$t_{15} = 1.91$ $p < 0.05$

The results indicate that significant click migration has taken place towards the putative boundaries. Different patterns of response can be observed between the two response conditions. This can be summarised by saying that boundary markers in Condition 1 become significant and the markers in Condition 2 become much more potent in the retrospective response condition, while the displacement created by Condition 3 remains the same under both response paradigms.

Discussion

The first observation to make is that there is a clear case for a non-verbal analogue for the click migration phenomenon in speech. The results obtained here indicate that extraneous noises are drawn towards boundary markers in musical stimuli. Musically unsophisticated listeners appear to segment non-verbal stimuli in much the same way as speech, using salient physical and structural features as boundary markers. Hence, the primary aim of this experiment, which was to demonstrate that non-verbal stimuli are processed in a functionally similar way to speech within this paradigm, has been realised. The core of this proposition is not that all systems involved in the perception of speech are interchangeable with those used to perceive other non-verbal sounds, but that all auditory events are tackled by qualitatively similar processes. There is a common set of features which can be observed in the processing of all auditory stimuli.

In the context of this experiment these common features can be seen by looking more closely at the results. The starting point for this is the different pattern of results obtained for the two response conditions. In Response A, only two conditions showed significant click migration - Condition 2 and Condition 3. Both of these conditions used physical markers to indicate boundary positions. Condition 1, which only used syntactic markers, did not elicit a significant degree of displacement. Clearly, when an immediate response is required, processes which are directed at physical features are dominant. This pattern can be contrasted with the results from Response B where there is strong evidence to indicate that subjects were also using structural markers to segment the stimulus. When the response is delayed until the end of the stimulus, Condition 1 demonstrates a highly significant level of migratory activity. This result indicates that subjects were influenced by purely syntactic boundary markers when the whole structure of the stimulus had been revealed.

There appear to be at least two processes at work in identifying segmental boundaries; one uses surface cues monitoring physical markers on a moment by moment basis, while the other integrates information into the final percept which is the result of a more extended interpretation of underlying structure. There is clear evidence that these two processes, although functioning independently, have a cumulative effect on the percept. This comes from Condition 2 where the degree of attraction of a boundary increases greatly from the immediate report condition, where only the physical markers appear to be salient, to the retrospective report condition where both physical and structural features combine to emphasise the boundary position.

The processes identified here seem to mirror the function of grammar and prosody in speech. Therefore, if musically unsophisticated subjects identify and utilise physical and structural segmental information from non-verbal stimuli in much the same way as they use grammatical and prosodic markers in speech perception, it follows that these are common features of qualitattively similar systems.

Experiment 3.ii

Introduction

The main aim of this experiment was to investigate whether clicks in music are percieved consistently earlier or later than their actual position. In speech reseach, clicks are perceived early. Sasaki (1980), who directly compared speech and music, reported that in both types of stimuli clicks were perceived earlier than their actual position. However, Gregory (1978) looked at migration in very brief musical fragments and observed that subjects appeared to perceive clicks later than their actual position. This contradiction was not really settled by Sloboda and Gregory (1980), whose results indicated that clicks were perceived slightly late but interpreted their findings as a consequence of onset asynchrony between the click and the note with which it coincided. Sloboda and Gregory recalibrated their results by the amount of the asynchrony and concluded that the clicks were really perceived earlier than their actual position.

Therefore, the sum of the existing evidence does not offer any decisive indication that clicks in musical stimuli behave in the same way as they do in speech. In order to provide some clear information on this subject, the present experiment was carried out. It followed the parameters of Experiment 3.i, with the exception that the clicks were placed at a putative boundary rather than before and after it.

Method

Design

This experiment followed exactly the same design parameters as Experiment 3.i. boundary definition and response indication being precisely reproduced. This gave rise to a three by two design, with boundary definition being a within-subjects variable and time of response a between-subjects variable. The only difference between this experiment and Experiment 3.i was that the click was placed at the boundary point, rather than before and after.

Stimulus Construction Preparation and Presentation

The same basic stimulus tokens were used in this experiment as were used in 3.i. However, instead of placing the extraneous clicks before or after the boundary point, they were inserted at the boundary. The purpose of placing the clicks at the boundary was to enable a clear measurement of the system's propensity to perceive the extraneous noise earlier or later than its actual position by eliminating the possibility of migratory interference.

Subjects

Subjects were all undergraduates at the City University, London. Their ages ranged from 18 years to 29 years with an average of 22 years. The group of 27 consisted of male and female subjects. All subjects were paid for their participation. Subjects were enlisted on the basis that they had not undergone any specialist musical training beyond ABRSM Grade 5. This stipulation was designed to exclude music specialists of all kinds. However, some musical training was allowed in order that the sample should reflect the level of knowledge in the population. Subjects were selected on the same criteria as in the previous studies presented here.

Procedure

The procedure followed in this study was exactly the same as the previous experiment. Subjects were tested in groups of not more than 10. The instructions informed subjects that they would be hearing a number of musical themes and they would have three tasks to perform after each musical presentation. Each theme was displayed as an overhead projection, (described above in Experiment 3.i - Response Indication). In the immediate response condition the projection of each stimulus was visually available during the whole trial to enable subjects to respond at once when the click occurred. Conversely, in the retrospective response condition the overhead representation of the theme was not displayed until the stimulus offset, thus making it impossible for subjects to respond until the end of the trial. Subjects were asked to give a number corresponding to the position of the extraneous noise and ratings on the other tasks. Subjects were told to divide their attention equally between the tasks.

Two practice trials were then given, after which subjects could ask questions. Every effort was made to ensure that all of the subjects understood the instructions. After each experimental trial, subjects made their responses on an answer sheet, shown in Appendix 3.

Results and Analysis

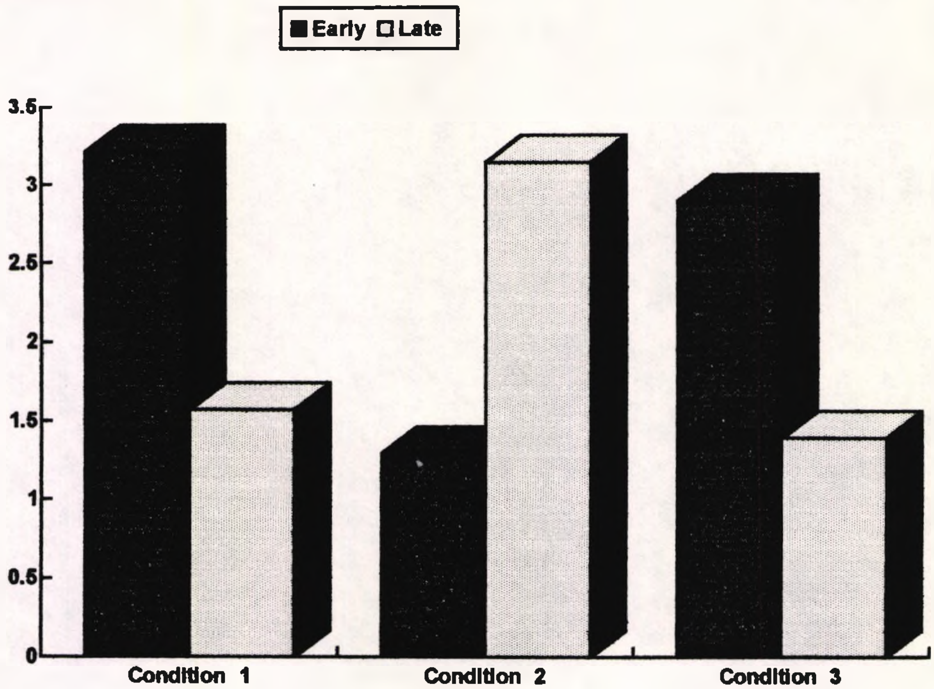
In order to examine the level of anticipation of the extraneous noise the judgements that the subjects made were divided into two categories: Those which placed the click prior to its actual position 'Early', and those which placed the click after the event 'Late'. Each judgement counted as one unit, Early or Late, no matter how far it was misplaced from its real position. Accurate judgements were counted as zero.

The data shown below in Table 8.3 and in Figure 8.5 depict the mean number of Early/Late judgements in each stimulus condition per subject.

Table 8.3

	Condition 1	**Condition 2**	**Condition 3**
Early	3.21	1.29	2.89
Late	1.57	3.14	1.39

Figure 8.5



The data shown in Table 8.3 and Figure 8.5 seem to indicate that in two of the three conditions, 1 and 3, the majority of inaccurate judgements placed the click earlier than its actual position, while the other condition, 2, more clicks were perceived late. These data were analysed using a three-way ANOVA. The analysis indicated that there was a significant difference between the number of clicks perceived early and those which were perceived late :

$$F(1,26) = 4.514 \quad p < 0.05$$

This result means that significantly more clicks were perceived before their actual position than after. This finding indicates that, as in speech, when an extraneous noise is mislocated there is a significant tendency to anticipate its real position.

However, an unexpected aspect of the findings was the interaction between the stimulus conditions and the judgments on the click position. This is clearly illustrated in Figure 8.6.

Figure 8.6

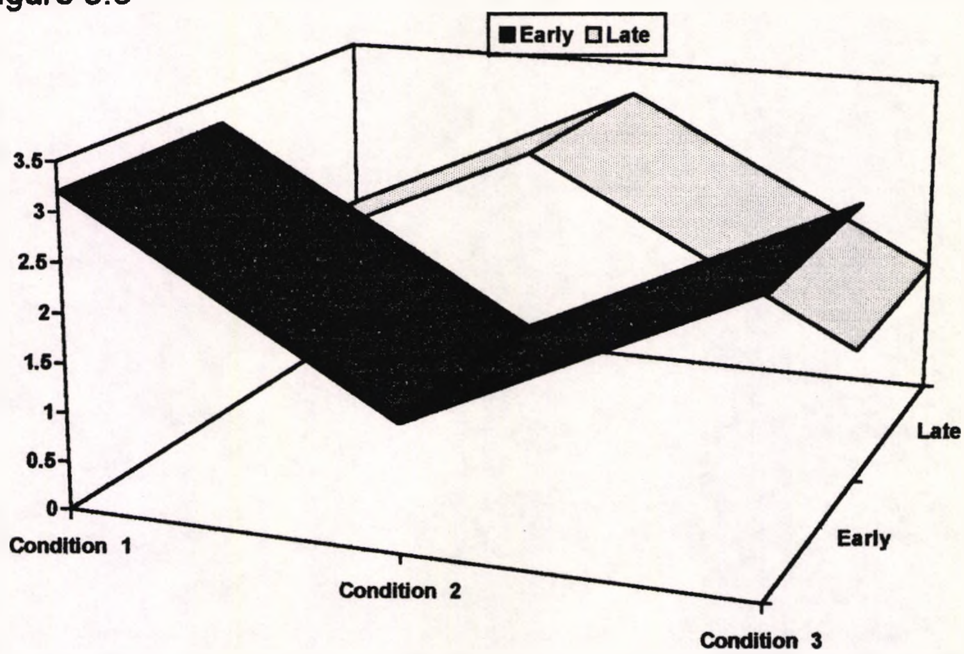


Figure 8.6 shows that in Conditions 1 and 3 the clicks are perceived prior to their actual position more often than they are perceived late. But in Condition 2 this state of affairs is reversed with more of the mislocated clicks perceived late. This interaction was highly significant :

$$F(2, 52) = 19.544 \quad p < 0.01$$

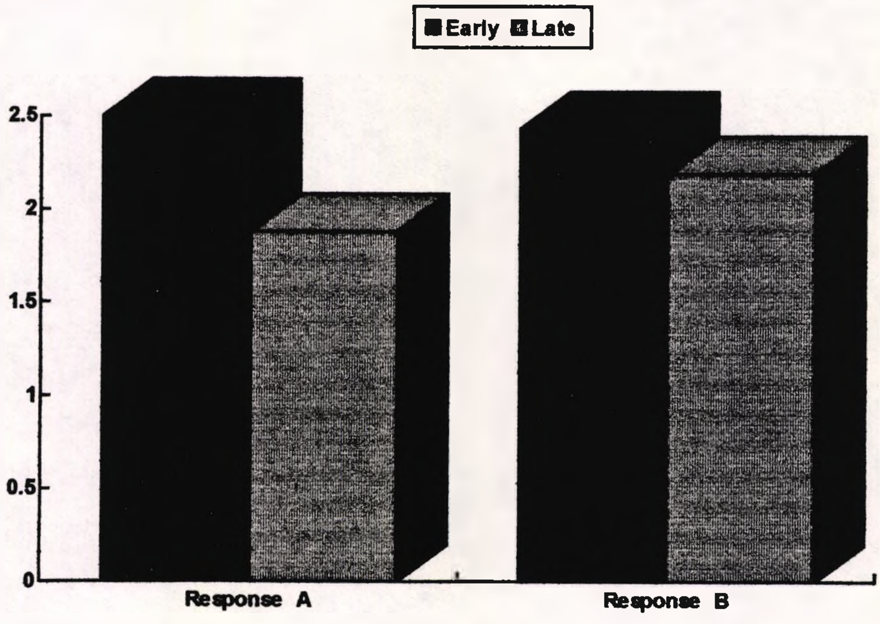
This interaction was not predicted by any aspect of this thesis or in the literature.

It was also part of the analysis to examine any possible difference in click anticipation caused by the two types of response protocol. Table 8.4 and Figure 8.7 show the number of early and late judgements made under the two response protocols.

Table 8.4

	Response A	Response B
Early	2.50	2.43
Late	1.88	2.19

Figure 8.7



The analysis of these data show that there is no significant interaction between the early or late mislocation of clicks and the protocol under which the response was made:

$$F(1,26) = 0.892 \quad p > 0.05.$$

This result indicates that it did not matter at what point during the stimulus presentation the response was made, subjects were still just as likely to anticipate the position of the extraneous noise.

Discussion

The results of this experiment show clear evidence for the proposition that when clicks are mislocated they are perceived earlier than their actual position in musical stimuli. This finding agrees with the conclusions of Sasaki (1980) and the 'adjusted' data of Sloboda and Gregory (1980). As clicks placed in speech stimuli are also perceived earlier than their actual position, these results give a further indication that the processes which analyse non-verbal auditory information are analogous to verbal processing systems.

The aspect of these findings which proved to be unexpectedly significant was the interaction of stimulus conditions with the position of the mislocated judgements. Condition 2 which combined structural and intonational boundary markers reversed the trend of the other two conditions, by eliciting more 'late' judgements than 'early' mislocations. It is not at all clear why this rather anomalous finding has been produced. Rather than try to theorise on the possible underlying explanation of this result without any converging evidence, it is perhaps more realistic to view it as an intriguing oddity.

Therefore, in conclusion it can only be reiterated that, in the absence of a more insightful explanation of the unexpected interaction, overall there is clear evidence for parallels in the pattern of click anticipation of verbal and non-verbal stimuli.

CHAPTER 9

General Discussion   Experimental Series 3

The findings of the experiments in this section represent a clear case for the existence of click migration in non-verbal stimuli. This case can be made on two grounds, methodological and theoretical.

The methods employed in this section were successful in two respects; they enabled musically unsophisticated subjects to make a judgement on the position of a click without the use of reaction time procedures, and the musical information was displayed without offering any visual cues to segmentation. These methodological conditions avoided some of the drawbacks of previous studies, and made it possible to use a representative sample of the population, rather than a restricted group of music specialists. While there is nothing wrong in using music specialists to answer more specific questions in music theory, testing a broad cross-section of the population gives the findings of a study a wider application.

The theoretical implications may be examined on three levels: as an analogue to click migration in speech, as limited information pertaining to the perception of musical structure, and as a provider of insight into the nature of auditory processing systems.

In terms of an analogue to click migration in speech, both Experiments 3.i and 3.ii produced evidence in support of this concept. Experiment 3.i demonstrated that the musical equivalents of grammatical and prosodic boundary markers in speech also attract extraneous clicks in the non-verbal stimuli. Furthermore, Experiment 3.ii showed clearly that, just as in speech, mislocated clicks in general are perceived earlier than their actual position. Taken together with the results from the other studies on click migration in music, these findings make a very strong argument for asserting that segmentation systems in non-verbal stimuli are analogous to those operating on speech sounds.

Experiment 3.i can also be seen as providing some limited information on the subject of perceived musical structure. Results from this experiment indicate that a boundary defined in purely structural terms attracted significant click migration. This demonstrated that subjects with little or no musical training were routinely processing musical stimuli in terms of its structural features. However, it seems that the whole structure has to be appreciated before a parsing decision can be made. As Experiment 3.i only examined structural features which were clarified on the completion of the stimulus, a further study might wish to elaborate the musical context, with the aim of establishing the structural boundary prior to the click. This would establish a more complete picture of the way purely structural information is used in an a priori as well as a post-hoc context.

The findings in this section also give an insight into the nature of the processes which are involved in segmentation. The data suggest that there are at least two different processes at work within the system, one which is sensitive to the surface characteristics of the ongoing acoustic input, and another which operates in a non-linear mode analysing a deeper level of information. The former process appears to use physical characteristics as a basis for segmentation decisions, while the latter attends to the structural elements of the array which might only be comprehended after the boundary position. This means that some features of the stimulus must undergo post hoc reintegration into the final percept.

CHAPTER 10

Conclusions and Overview

This thesis began with two theoretical aims. The primary aim was to examine the extent and the nature of the analogue in the non-verbal auditory world for the well established speech based illusions of transformation, restoration and migration. The secondary aim was to assess the subsequent experimental findings and examine their implications for auditory processing. Hence, these conclusions will be divided in two, each part concentrating on a separate theoretical aim.

Illusions in Speech and their Non-Verbal Analogues

The various experiments presented in this thesis have all shared one aspect of their purpose, namely, to provide evidence of non-verbal analogues for illusory phenomena previously observed in speech. In order that the implications can be more comprehensively evaluated, it might be of some use to the reader to offer a brief recap of the theoretical basis of each analogue and the relevant general findings from each section.

The first speech based illusion to be tackled in this thesis was the verbal transformation effect. The findings of the experiments presented here demonstrated that subjects perceived illusory changes in stabilised non-verbal stimuli. The basis of the analogue in this case was the quality of the transformations perceived by the listeners. This stems from the contention of Warren (1982), that non-verbal stimuli could not elicit transformations which so radically distorted the percept as to change its identity, because the cognitive mechanisms which process auditory input are significantly less creative than those which function in speech perception. Warren concluded that there would be a qualitative gap between any transformations reported for non-verbal sounds and those reported for speech.

Therefore, it is the quality of the illusory changes generated by verbal and non-verbal stimuli that lies at the heart of any proposed analogue. The means of assessing the qualitative equivalence of transformations elicited by different types of auditory input was based on the concept of stimulus identity. This meant that illusory changes which distorted the essential nature of a repeating token could be equated with any other transformation which displayed a similar level of disparity from the original stimulus.

The results indicated that music and other complex sounds elicited as many illusory changes in identity and thus a genuine analogue was clearly demonstrated. Subjects experienced strong illusory changes in all stimulus conditions, dispelling the notion that transformations generated by stabilised auditory input are limited to the verbal domain.

The attempt to demonstrate a non-verbal analouge for the phonemic restoration effect did not rest on a valid means of comparing the different types of auditory input, but instead on the hypothesis that the continuity illusion in non-verbal stimuli, functions, as in speech, by using concept-driven as well as data-driven aspects of the token. This basis for an analogue arose from the historical view that the continuity illusion in speech was led by concept-driven factors while the same phenomenon in non-verbal stimuli was entirely mediated by data-driven factors.

The theory proposed by this thesis was that if higher level variables were observed to influence the restoration of missing segments of non-verbal stimuli then it would be difficult to draw a functional distinction between the way that continuity is established in different auditory domains.

However, the results obtained from examining restoration of missing segments from real musical themes failed to successfully disentangle the effects of the variables under consideration. Further studies, using more specifically designed stimulus material, did reveal significant effects of concept-driven variables. Hence if higher level factors are influential in the restoration of missing segments of both speech and non-verbal stimuli then the processes which underlie the maintenance of continuity in both domains can be considered analogous.

The final illusory phenomenon examined in this thesis was the click migration effect. The basis of the non-verbal analogue within this paradigm was that in speech extraneous noises tended to be perceptually attracted towards segmental boundaries, making use of both structural and prosodic markers.

In speech the structural markers correspond to grammatical features of the utterance, while the prosodic markers are realised by the surface intonational pattern of the stimulus. It appears that these markers are used by the system to segment the auditory stream into perceptual units. In order to create a non-verbal analogue, structural and intonational aspects of a musical stimulus were manipulated to reproduce the features found to be salient in speech.

The findings demonstrated that the extraneous noises placed in the musical stimuli followed the same pattern of mislocation observed in speech. The clicks migrated towards boundaries that were marked by structural features and /or surface intonation. Under these criteria it can be asserted that at least musical stimuli - for other non-verbal sounds were not explored in these experiments - perform in an analogous way to speech within the click migration paradigm.

In all the illusory phenomena examined by this thesis clear evidence has emerged that non-verbal stimuli perform in a functionally similar way to speech. Yet it has been the routine practice in the literature to cite these phenomena as part of the case for viewing speech as a unique form of auditory perception. It was asserted that these illusions were examples of the singular way in which speech perception was not shackled by the same forces which ruled over other auditory domains.

The strong interpretation of these findings is that speech should no longer be regarded as a unique form of auditory processing, as the evidence gleaned from the examination of three phenomena has converged on the same conclusion that music and other complex sounds are capable of producing comparable illusory effects. However, a more considered appreciation of the overall position would concede that the theory of the special nature of speech perception stands on more than one evidential pillar. The arguments from the standpoints of linguistics and neuropsychology remain untouched. So what is the most balanced interpretation of the analogies demonstrated by this thesis?

At a general level there is still a great deal of evidence to support the 'speech is special ' argument. For example, there is no need to question whether listeners learn language in a unique way, or the localisation of the speech centres in the brain. Nothing presented here reflects directly these concepts. These aspects of speech perception may indeed be special. However, this thesis does not contend that all aspects of speech perception are interchangeable with all aspects of every other auditory stimulus. That would be futile.

It is the contention here that the processes which sort and sift all auditory sensory data operate on the same general principles. This does not mean that the processes which convert the speech stream into internal representations are less creative than was previously envisaged, but that all auditory events undergo similar qualitative processing. Furthermore, it should be noted that although the same sort of processing is assumed, there is no intention to imply that the same 'processors' are involved. That is, there is no argument that the physical identity of the processors are the same, just that the processes involved function under similar constraints.

The findings obtained here show that non-verbal stimuli are segmented and restored in the same fashion as speech and that illusory transformations occur across a wide range of stabilised tokens. These results indicate that the nature of the processes which control auditory perception have more in common than was previously theorised. Therefore, it can be confidently asserted that although the perception of speech may not be completely interchangeable with other complex sounds in processing terms, it is not entirely unique.

The Processing Implications of Non-Verbal Illusions

In demonstrating non-verbal analogues for illusions previously thought only to exist in speech perception the work presented here has come to the conclusion that there must be some essential similarities in the processing of different auditory stimuli. The points of interest raised by the examination of each phenomenon will be considered here in turn.

The discovery of strong illusory transformations in auditory domains outside speech required the most profound reassessment of the theoretical basis of all the phenomena examined by this thesis. The most compelling theory which had been proposed to explain the effect centred on the concept that the illusions experienced by subjects were the result of the exceptionally creative processes used by the perceptual system to interpolate spoken input.

However, if the above theory were correct then no significant illusory changes would be reported when non-verbal stimuli were used as repeating tokens. The fact that a non-verbal analogue was established comprehensively refuted any theory which was based on the premise of uniquely creative speech processing mechanisms. Furthermore this thesis took the view that it was essentially illogical to attribute the effects of this phenomenon, which often distort the original percept beyond recognition, to processes which must be dedicated to creating a reliable representation of external events.

In contrast, this thesis chose to completely re-evaluate the nature of the phenomenon. Instead of viewing illusory changes of percept as a reflection of the creative aspect of normal processing, these transformations were seen as the direct result of the abnormal way in which the perceptual system functions under the pressure of a repeating token. Under this hypothesis, representations which are related to the stimulus are accessed continuously, causing some of them to be perceived at a conscious level.

This concept was called the Associative theory. It was based on the notion that when the system perceives any token a web of interconnected internal representations is stimulated. Hence, when any item is presented to the system the effect is not limited to the accessing of that item, but also acts as a preparatory stimulus for other interrelated items. In everyday circumstances this process of associative priming is an adaptive mechanism speeding perception. However, in the stabilised auditory paradigm this process is subverted. The repeated presentation of a token causes connected representations to receive inappropriately continuous stimulation, eventually accessing them as percepts.

In general terms it might be tempting to interpret the Associative theory's explanation of illusory transformations as support for a connectionist description of auditory processing, for instance McLelland and Kawamoto (1986). This may seem to be the case, as the illusory percepts appear to be the result of inappropriate levels of activation within an interconnected system. However, although the Associative theory presupposes a complex web of interrelated internal representations it does not purport to test any form of connectionist hypothesis. It might be possible to construct a series of auditory transformation experiments which specifically explore connectionist hypotheses in processing, but this must remain outside of the scope of the present enquiry.

In assigning the cause of illusory transformations to the activation reverberating through associated representations this phenomenon links with the second area of study, the restoration of missing segments. Part of the theoretical framework which was formulated to facilitate research on the continuity illusion in complex non-verbal sounds incorporated the concept of directed activation of related representations.

This framework (see Figure 10.1) postulates that at the offset of A the processing system activates a range of possible stimulus completion alternatives. This hypothesis was supported in Experiment 1.ii where the illusory changes elicited under the stabilised auditory stimulus paradigm were focused around the missing segment of the repeating token, thereby indicating that the activation of associated representations was intensified by the processing of the ambiguous segment.

Figure 10.1

T------------------>

[A______N/✠______B]

P

A = The auditory events preceding N. T = Time.

B = The auditory events following N. P = Whole percept.

N = Replacement noise/silence. ✠ = Missing/restored element.

The prior activation of stimulus-related representations was also demonstrated in Experiments 1.iii and 2.iv, by different priming regimes. In the former more illusory changes were reported, while in the latter restoration was increased. These findings offer general support for the concept that the level of activation in the internal representations related to the stimulus token plays an important role in both phenomena. In the case of illusory transformations the nature of the experimental paradigm causes the associative activation to disrupt and subvert the percept, while similar processes examined under the continuity format, which more accurately reflects a normal listening situation, are adaptive and enhance the percept.

The findings also emphasised the effect of melodic predictability on the perceived restoration of the missing segment. Predictable themes received higher continuity ratings than unpredictable themes. This result appeared to demonstrate that concept-driven variables as well as data-driven variables could influence the processes which control restoration.

However, these findings seemed to directly contradict the results obtained by DeWitt and Samuel (1990). They found that high predictability had a negative effect on perceived continuity. The interpretation which Dewitt and Samuel offered was that their findings were in line with the modularity theories of Fodor (1983). They claimed that the increased restoration found in missing notes within chords and scales and the concomitant decrease in the restoration of highly predictable stimuli, reflected Fodor's distinction between "low level / entrenched" and "high level / constructed" representations. DeWitt and Samuel suggested that the "entrenched" representations, scales and chords, were within the module and thus "play a role in perception", while the "constructed" representations, as exemplified by predictability, do not.

It is not the purpose of this thesis to offer any general opinion on the validity of Fodor's modularity theory, but it is within the scope of this discussion to examine DeWitt and Samuel's interpretation of their results. The role of the concept-driven variable predictability appears to be key. The evidence provided in this thesis shows predictability as a positive influence on restoration, while DeWitt and Samuel found a negative effect. The interpretation of these findings which makes the most coherent sense of the apparently divergent results involves viewing predictability as a variable which can affect continuity at any point along a continuum of influence.

It is possible to imagine circumstances under which elements of a stimulus are so unambiguously predictive that they will reduce the likelihood that the system will accept any representation of the expected event which is less than perfect. Similarly, an event which is only implictly predicted could prepare the system without erecting a barrier to the acceptance of an imperfect representation of the expected event.

This interpretation sees the perceptual system using predictability information, in restoration, on a case-by-case basis, interpolating the level of expectation generated with the amount of data-driven conformation provided by the stimulus. Hence under this hypothesis it is perfectly congruent for experiments on predictability to show positive or negative influence on continuity. The critical feature of the finding is not the direction of the result, as this would hinge on the content of the manipulation, but whether the variable has a significant effect at all.

The view of the role of predictability as expressed above is part of a general perspective on restoration, and for that matter click migration, which sees the phenomena as adaptive analytical auditory processes applicable across stimulus domains, rather than as forms of proof for any particular theory - modularity or connectionism. In a sense the position taken here is that it is more important to observe the action of the processing system clearly through the established phenomena, than to attempt to fit results into any existing set of theoretical constructs.

In order to expand this concept further it is necessary to return to the experimental findings. One aspect of the findings which emerged from both the continuity studies and click migration was that information occurring after the critical point in a stimulus could affect its processing in terms of the final percept. In the continuity studies it was shown that repeating the context of a missing segment at a point later in the stimulus, increased restoration ratings. It was also revealed that the nature of the critical note context was relevant to the processing system, as some note contexts were more readily assimilated into the final percept than others. This aspect of the findings indicated that the processes involved in restoring missing elements were operating discriminatory assessments of the information occurring after the critical segment. In a sense, they seem to be calculating the value of the post-hoc data for the missing element.

The process of post-hoc percept adjustment can be seen to be even more sophisticated in the light of findings from the click migration studies structural boundary markers could be seen to attract extraneous noises, only when the complete extended structure of the stimulus had been assimilated. This demonstrated that both restorative and segmental processes are actively engaged in reorganising the nature of the percept on the basis of concept-driven information occurring after an ambiguous event or clouded segment.

The other major finding from the area of click migration was that the systems which control segmentation not only use structural markers but surface markers. This is a further example of the flexibility of the processing system, demonstrating that a range of information is available for segmenting a complex auditory stream.

The wider conclusion which can be drawn from the findings of this thesis is that the processes which underlie the construction of the internal representations of non-speech stimuli, as demonstrated by the findings presented here, are rich, pro-active and elegant. After considering the evidence from the three illusory phenomena presented here it is impossible to conclude that speech is a uniquely creative form of auditory processing as music and a range of complex sounds from the everyday world undergo similar varieties of synthesis.

REFERENCES

Bever, T.G., Lackner, J. & Kirk, R. (1969) The underlying structures of sentences are the primary units of speech processing. Perception & Psychophysics, 5, 225-234.

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

Bregman, A.S. & Dannenbring, G.L. (1977) Auditory continuity and amplitude edges.Canadian Journal of Psychology, 31, 151-159.

Calef, R.S. Calef, R.A.B. Kesecker, M.P. & Burwell, R. (1974) Verbal transformations of stabilised taboo and neutral words. Perceptual and motor skills, 38, 177-178.

Carroll, J.M. & Bever, T.G. (1976) Sentence comprehension: A case study in the relation of knowledge to perception. In Carterette, E.C. & Friedman, M.P. (Eds), The Handbook of Perception. Vol. 7: Language and Speech. New York: Academic Press.

Carroll, J.M. & Tanenhaus, M.K. (1975) Functional clauses and sentences segmentation. Unpublished manuscript.

Chapin, P.G., Smith, T.S. & Abrahamson, A.A. (1972) Two factors in perceptual segmentation of speech. Journal of Verbal Learning and Verbal Behaviour. 11, 164-173.

Ciocca, V. & Bregman, A.S. (1987) Perceived continuity of gliding and steady-state tones through interrupting noise. Perception & Psychophysics, 42, 476-484.

Dannenbring, G.L. (1976) Perceived auditory continuity with alternately rising and falling frequency transitions. Canadian Journal of Psychology, 30, 99-114.

DeWitt, L.A. & Samuel, A.G. (1990) The role of knowledge-based expectations in music perception: Evidence from musical restoration. Journal of Experimental Psychology: General, 2, 123-144.

Ditchburn, R.W. & Fender, D.H. (1955) The stabilized retinal image. Optica Actor, 2, 128-133.

Elfner, L.F. (1969) Continuity in alternately sounded tone and noise signals in a free field. Journal of the Acoustical Society of America, 46, 914-917.

Elfner, L.F. (1971) Continuity in alternately sounded tonal signals in a free field. Journal of the Acoustical Society of America, 49, 447-449.

Elfner, L.F. & Caskey, W.E. (1965) Continuity in alternately sounded tone and noise signals as a function of manner of presentation.Journal of the Acoustical Society of America, 38, 543-547.

Elfner, L.F. & Homick, J.L. (1966) Some factors affecting the perception of continuity in alternately sounded tone and noise signals. Journal of

Elliot, L.L. (1963) Apprent change of repetative noise bursts. Journal of the Acoustical Society of America, 35, 1917-1973.

Fenelon, B. & Blayden, J.A. (1968) Stability of auditory perception of words and pure tones in under repetative stimulation in neutral and suggestibility conditions. Psychon. Sci., 285-286.

Flores d'Arcais, G.B. & Schreuder, R. (1983) The process of language understanding: A few issues in contemporary psycholinguistics. In Flores d'Arcais, G.B. & Jarvella, R.J. The Process of Language Understanding. Wiley.

Fodor, J.A. (!983) The Modularity of Mind: An Essay on Faculty Psychology. M.I.T. Press.

Fodor, J.A. & Bever, T.G. (1965) The psychological reality of linguistic segments. Journal of Verbal Learning and Verbal Behaviour, 4, 414-420.

Fodor, J.A., Bever, T.G. & Garrett, M.F. (1974) The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. New York. McGraw-Hill.

Foss, P.J. & Blank, M.A. (1980) Identifying the speech codes. Cognitive Psychology, 12, 1-31.

Garrett, M.F., Bever, T.G. & Fodor, J.A. (1966) The active use of grammar in speech perception. Perception & Psychophysics, 1, 30-32.

Gregory, A.H. (1978) Perception of clicks in music. Perception & Psychophysics, 24, 171-174.

Handel, S. (1989) Listening. The M.I.T Press.

Houtgast, T. (1972) Psychophysical evidence for lateral inhibition in hearing. Journal of the Acoustical Society of America, 51, 1885-1895.

Houtgast, T. (1973) Psychophysical experiments on "tuning curves" and "two- tone inhibition." Acustica, 29, 168-179.

Houtgast, T. (1974) Masking patterns and lateral inhibition. In E. Zwicker and E. Terhardt (Eds), Facts and models in hearing. Berlin: Springer-Verlag, 258-265.(b)

Houtgast, T. (1974) The slopes of the masking patterns . In E. Zwicker and E.Terhardt (Eds), Facts and models in hearing. Berlin: Springer-Verlag, 269-272.(c)

Ladefoged, P. & Broadbent, D.E. (1960) Perception of sequence in auditory events. Quarterly Journal of Experimental Psychology. 12, 162-170.

Lass, N.J. & Golden, S.S. (1971) The use of isolated vowel sounds as auditory stimuli in elicting the verbal transformation effect. The Canadian Journal of Psychology. 25, 349-359.

Lass, N.J., West, L.K. & Taft, D.D. (1973) A non-verbal analogue to the verbal transformation effect. 27 (3), 273-279.

Layton, B. (1975) Differential effects of two nonspeech sounds on phonemic restoration. Bulletin of the Psychomic Society, 6, 487-490.

Lerdahl, F. & Jackendoff, R. (1983) A generative theory of tonal music. The M.I.T Press.

Liberman, A.M. (1982) On finding that speech is special. American Psychologist, 2, 148-167.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967) Perception of the speech code. Psychological Review, 74, 431-461.

Meyer, D.E. & Schvaneveldt, R. W. (1971) Facilitation in recognizing pairs of words : evidence of dependence between retrieval operations. Journal of Experimental Psychology, 90, 227-234.

McAdams, S. & Bregman, A.S. (1979) Hearing musical streams. Computer Music Journal, 3, 26-43.

McCarthy, R.A. & Warrington, E.K. (1990) Cognitive Neuropsychology, Academic Press.

McGurk, H. & MacDonald, J. (1976) Hearing lips and seeing voices. Nature, 264, 746-748.

McLelland, J.L. & Kawamoto (1986) In; Parallel distributed processing : Explorations in the microstructure of cognition. Rumelhart, D.E. & McLelland, J.L. (eds). The M.I.T Press.

Miller, G.A. & Licklider, J.C.R. (1950) The intelligibility of interrupted speech. Journal of the Acoustical Society of America, 22, 167-173.

Obusek, C.J. & Warren, R.M. (1973) Relation of the verbal transformation and the phonemic restoration effects. Cognitive Psychology, 5, 97-107.

Ojemann, G.A. (1982) Inter-relationships in the brain of language-related behaviours: Electrical stimulation mapping. In Kirk, U. (Ed) The neuropsychology of language, reading and spelling. Academic Press.

Perl, N.T. (1970) The application of the verbal transformation effect to the study of cerebral dominance. Neuropsychologica, 8, 259-261.

Samuel, A.G. (1981) The role of bottom up confirmation in the phonemic restoration illusion. Journal of Experimental Psychology: Human perception and performance, 7, 1124-1131.(a)

Samuel, A.G. (1981) Phonemic restoration: Insights from a new methodology. Journal of Experimental Psychology: General, 110, 474-494.(b)

Sasaki, T. (1980) Sound restoration and temporal localisation of noise in speech and music sounds. Tohuku Psychologica Folia, 39, 79-88.

Severance, E. & Wasburn, M.F. (1907) Minor studies from the psychological laboratory of Vassar College.iv. 18, 182-186.

Skinner, B.F. (1936) The verbal summator and a method for the study of latent speech. Journal of Psychology, 2, 71-107.

Sherman, G.L. (1971) The phonemic restoration effect : An insight into the mechanisms of speech perception. Unpublished M.S. dissertation, University of Wisconsin-Milwaukee.

Sloboda, J.A. & Gregory, A.H. (1980) The psychological reality of musical segments. Canadian Journal of Psychology, 34, 274-280.

Stoffer, T.H. (1985) Representation of phrase structure in the perception of music. Music Perception, 2, 191-220.

Swales, F. & Evans, C.R. (1967) V.T's and the stabilised retinal image. National Physical Lab, Autonomic Divn Report 37.

Taylor, M.M. & Henning, C.B. (1963) Transformations of perception with prolonged observation. Canadian Journal of Psychology, 17, 349-360.

Thurlow, W.R. (1957) An auditory figure-ground effect. American Journal of Psychology, 70, 653-654.

Thurlow, W.R. & Elfner, L.F. (1959) Continuity effects with alternately sounding tones. Journal of the Acoustical Society of America, 31, 1337-1339.

Thurlow, W.R. & Erchul, W.P. (1978) Understanding continuity effects with complex stimuli. Journal of the American Auditory Society. 4, 113-116.

Thurlow, W.R. & Marten, A.E. (1962) Perception of steady and intermittent sound with alternating noise-burst stimuli. Journal of the Acoustical Society of America, 34, 1853-1858.

Titchener, E.B. (1909) Lectures on the experimental psychology of the thought process. New York: Macmillan.

Vicario, G. (1960) L'effetto tunnel acustico. Rivista di Psicologia, 54, 41-51.

Warren, R.M. (1961) Illusory changes of distinct speech upon repetition - The verbal transformation effect. British Journal of Psychology, 52, 249-258. (a).

Warren, R.M. (1961) Illusory changes in repeated words: Differences between young adults and the aged. 74, 506-516. (b)

Warren, R.M. (1968) Verbal transformation effect and auditory perceptual mechanisms. Psychological Bulletin, 261-270. (c).

Warren, R.M. (1970) Perceptual restoration of missing speech sounds. Science. 167. 392-393. (b)

Warren, R.M. (1981) Perceptual transformations in vision and hearing. International Journal of Man-Machine Studies, 14, 123-132. (b).

Warren, R.M. (1982) Auditory Perception: A New Synthesis. New York: Pergamon.

Warren, R.M. & Gregory, R.L. (1958) An auditory analogue of the visual reversible figure. American Journal of Psychology. 71, 612-613.

Warren, R.M. & Obusek, C.J. (1971) Speech perception and phonemic restorations. Perception & Psychophysics. 9. 358-362.

Warren, R.M. Obusek, C.J. & Ackroff, J.M. (1972) Auditory induction: Perceptual synthesis of absent sounds. Science, 176, 1149-1151.

Warren, R.M. & Sherman, G.L. (1974) Phonemic restorations based on subsequent context. Perception & Psychophysics, 16, 150-156.

Warren, R.M. & Warren, R.P.A. (1966) A comparison of speech perception in childhood, maturity, and old age by means of the verbal transformation effect. Journal of Verbal Learning and Verbal Behaviour, 5, 142-146.

Wiley, R.L. (1968) Speech communication using the strongly voiced components only. Doctoral dissertation, Imperial College, University of London.

Wingfield, A. & Klein, J.F. (1971) Syntactic structure and acoustic pattern in speech perception. Perception & Psychophysics, 9, (1A) 23-25.

APPENDIX 1

Categories of Identity Change

Global change

Missing section

Extra section

Within stimulus pitch change

Within stimulus separation

Figure-ground

Streaming

Additional familiar sound

Rhythm

Categories of Non-Identity Changes

Emphasis

Volume

Rate

Clarity

Between stimulus pitch change

Other

Semantic loss

List of themes used in Experiment 1.ii

Familiar Themes

Sousa, Liberty Bell March 1st Theme

Prokofieff, Peter and the Wolf Op 67 1st Theme

Tschaikovsky, Swan Lake Op 20a 1st Movement Intro

Grieg, Peer Gynt Op 46 1st Movement Morning Mood

Unfamiliar Themes

Sousa, King Cotton March 2nd Theme

Prokofieff, Sonata in D Op 94 4th Movement 3rd Theme

Tschaikovsky, Serenade in C Op 48 4th Movement 1rd Theme

Grieg, French Serenade Op 62

List of complex sounds used in Experiment 1.iii

Sea & Wind (practice)

Photocopier

Heartbeat

Water pouring

Gibbon Cries

Frogs Croaking

Filing Metal

APPENDIX 2

Table A.Total Continuity Ratings and Percentage Restoration for each Theme in Experiment 2.i

	Familiar			
Theme	Strong		Weak	
1	110	56.1%	115	58.7%
2	153	78.1%	141	72%
3	110	56.1%	154	78.6%
4	97	49.5%	104	53.1%
5	119	60.7%	160	81.6%
6	124	63.3%	113	57.7%
7	100	51%	134	68.4%
8	123	62.8%	143	73%
9	140	71.4%	130	66.3%
10	118	60.2%	168	85.7%
11	102	52%	116	59.2%
12	105	53.6%	105	53.6%
13	115	58.7%	109	55.6%
14	123	62.8%	135	68.9%
15	109	55.6%	121	61.7%

	Unfamiliar			
Theme	Strong		Weak	
1	52	26.5%	169	86.2%
2	102	52%	134	68.4%
3	102	52%	112	57.8%
4	153	78.1%	129	65.8%
5	136	69.4%	95	48.5%
6	139	70.9%	143	73%
7	134	68.4%	66	33.7%
8	108	55.1%	125	63.8%
9	117	59.7%	77	39.3%
10	116	85.7%	126	64.3%
11	113	57.7%	152	77.6%
12	149	76%	148	75.5%
13	127	64.8%	149	76%
14	144	73.5%	129	65.8%
15	102	52%	82	41.8%

List of Themes used in Experiment 2.i

Familiar Themes

1) Tschaikovsky, 1812, Festival Overture 2nd Theme

2) Ravel, Bolero Theme A

3) Beethoven, Symphony no. 6 in F Op. 68 1st Movement 1st Theme

4) Beethoven, Symphony no. 9 in D minor Op. 125 4th Movement 1st Theme

5) Beethoven, Für Elise.

6) Bach, Jesu, Joy of Man's Desiring 1st Movement 1st Theme

7) Mozart, Rondo a la Turca

8) Offenbach, Orpheus in Hades Act 1 Duo 2nd Theme

9) Prokofieff, Peter and the Wolf 1st Theme Peter

10) Bizet, Carmen Prelude to Act 1 1st Theme

11) Schubert, Symphony no. 8 in B minor 'Unfinished' 1st Movement 2nd Theme

12) Strauss Du and Du Waltzes from Die Fledermaus Op. 367 No. 1 2nd Theme

13) Prokofieff, Lieutenant Kije Op. 60 4th Movement Troika

14) Grieg, Peer Gynt Suite No. 1 4th Movement, In the Hall of the Mountain King

15) Grieg, Peer Gynt Suite 1st Movement, Morning Mood

Unfamiliar Themes

1) Tschaikovsky, Symphony No. 4 in F minor Op. 36 4th Movement 3rd Theme

2) Rameau, Suite in E minor, Rigaudon No. 1

3) Beethoven, Quintet in E flat Op. 16 3rd Movement Rondo

4) Beethoven, Serenade Op. 8 6th Movement

5) Beethoven, Sonata No. 3 in A Op. 69 1st Movement

6) Bach, Partita No. 6 in E minor Air

7) Mozart, Sonata in D flat K 570 1st Movement 2nd Theme

8) Offenbach, La Vie Parisienne Act V Entr'acte 1st Theme

9) Prokofieff, Alexander Nersky Op. 78 7th Movement

10) Bizet, Symphony No. 1 in C 3rd Movement

11) Schubert, Quartets No. 4 in C 4th Movement

12) Strauss Perpetuum Mobile Op. 257 Orch. Variation

13) Prokofieff, Gavotte Op. 12 No. 2

14) Grieg, Cradle Song Op. 68

15) Grieg, Norwegian Melody Op. 12 1st Theme

Circle One

a) Is the sound louder or quieter? L Q

No / Yes

b) Did the sound continue through the crackle? 1 2 3 4 5 6 7

c) How long, in seconds, do you think the sound lasted? ________ seconds

Circle One

a) Is the sound louder or quieter? L Q

No / Yes

b) Did the sound continue through the crackle? 1 2 3 4 5 6 7

c) How long, in seconds, do you think the sound lasted? ________ seconds

Circle One

a) Is the sound louder or quieter? L Q

No / Yes

b) Did the sound continue through the crackle? 1 2 3 4 5 6 7

c) How long, in seconds, do you think the sound lasted? ________ seconds

Circle One

a) Is the sound louder or quieter? L Q

No / Yes

b) Did the sound continue through the crackle? 1 2 3 4 5 6 7

c) How long, in seconds, do you think the sound lasted? ________ seconds

Tune:

Music did not continue Music continued

1 2 3 4 5 6 7

....

Were there two identical consecutive notes? YES NO

....

Subjective duration of tune:

0 1 2 3 4 5 6 7 8 9 10 seconds

*

Tune:

Music did not continue Music continued

1 2 3 4 5 6 7

....

Were there two identical consecutive notes? YES NO

....

Subjective duration of tune:

0 1 2 3 4 5 6 7 8 9 10 seconds

*

Tune:

Music did not continue Music continued

1 2 3 4 5 6 7

....

Were there two identical consecutive notes? YES NO

....

Subjective duration of tune:

0 1 2 3 4 5 6 7 8 9 10 seconds

*

Tune:

Music did not continue Music continued

1 2 3 4 5 6 7

....

Were there two identical consecutive notes? YES NO

....

Subjective duration of tune:

0 1 2 3 4 5 6 7 8 9 10 seconds

*

APPENDIX 3

Sex:

Age:

Musical Experience

Tune	Box No.							
		Musical	1	2	3	4	5	Unmusical
		Pleasant	1	2	3	4	5	Unpleasant
		Musical	1	2	3	4	5	Unmusical
		Active	1	2	3	4	5	Inactive
		Musical	1	2	3	4	5	Unmusical
		Interesting	1	2	3	4	5	Boring
		Musical	1	2	3	4	5	Unmusical
		Happy	1	2	3	4	5	Sad
		Musical	1	2	3	4	5	Unmusical
		Happy	1	2	3	4	5	Sad
		Musical	1	2	3	4	5	Unmusical
		Pleasant	1	2	3	4	5	Unpleasant
		Musical	1	2	3	4	5	Unmusical
		Active	1	2	3	4	5	Inactive
		Musical	1	2	3	4	5	Unmusical
		Interesting	1	2	3	4	5	Boring
		Musical	1	2	3	4	5	Unmusical
		Interesting	1	2	3	4	5	Boring
		Musical	1	2	3	4	5	Unmusical
		Happy	1	2	3	4	5	Sad
		Musical	1	2	3	4	5	Unmusical
		Active	1	2	3	4	5	Inactive
		Musical	1	2	3	4	5	Unmusical
		Happy	1	2	3	4	5	Sad
		Musical	1	2	3	4	5	Unmusical
		Pleasant	1	2	3	4	5	Unpleasant
		Musical	1	2	3	4	5	Unmusical
		Interesting	1	2	3	4	5	Boring
		Musical	1	2	3	4	5	Unmusical
		Active	1	2	3	4	5	Inactive
		Musical	1	2	3	4	5	Unmusical
		Active	1	2	3	4	5	Inactive
		Musical	1	2	3	4	5	Unmusical
		Interesting	1	2	3	4	5	Boring
		Musical	1	2	3	4	5	Unmusical
		Pleasant	1	2	3	4	5	Unpleasant
		Musical	1	2	3	4	5	Unmusical
		Happy	1	2	3	4	5	Sad
		Musical	1	2	3	4	5	Unmusical
		Happy	1	2	3	4	5	Sad
		Musical	1	2	3	4	5	Unmusical
		Interesting	1	2	3	4	5	Boring
		Musical	1	2	3	4	5	Unmusical
		Active	1	2	3	4	5	Inactive
		Musical	1	2	3	4	5	Unmusical
		Happy	1	2	3	4	5	Sad
		Musical	1	2	3	4	5	Unmusical
		Pleasant	1	2	3	4	5	Unpleasant
		Musical	1	2	3	4	5	Unmusical
		Active	1	2	3	4	5	Inactive

List of complex sounds used in Experiment 2.iv.

1. Goat
2. Washing machine
3. Milk bottles
4. Crackling flames
5. Tube train
6. Loo flushing
7. Filing cabinet drawer
8. Lemur cries
9. Seal cries
10. Photocopier
11. Lift door closing
12. Telephone ringing
13. Door being unlocked
14. Pages of book being turned
15. Jumping on floor
16. Steel door being opened and closed
17. Sash window
18. Clapping
19. Eating an apple
20. Rattling pills in a bottle
21. Vacuum cleaner
22. Cash register
23. Dialling on a telephone
24. Bird song