LUND UNIVERSITY

CENTRE FOR LANGUAGES AND LITERATURE

EFFECTS OF SUPRASEGMENTAL FEATURES ON THE PROCESSING OF SPOKEN WORDS IN THE HUMAN BRAIN : EVIDENCE FROM MISMATCH NEGATIVITY (MMN)

By Hatice Zora

Supervisor: Mikael Roll

Co-supervisor: Merle Horne

Master Thesis (2 Years) – SPVR01

Submitted to the graduate degree program in English Language and Linguistics as fulfillment of the requirements for the degree of Master of Arts

Lund University

2011 - Spring

Acknowledgements

This thesis would not have been possible without my supervisors Dr. Mikael Roll and Prof. Merle Horne, and their invaluable assistance, guidance and most importantly, their long-term patience with my progress. I hope to someday become half as good a supervisor as they have been.

I would also like to thank Prof. Carita Paradis, for her unconditional support. I am grateful to Dr. Yury Shtyrov at the MRC Cognition and Brain Sciences Unit for his helpful comments regarding experimental design and analysis. My gratitude also goes to Susanne Schötz for her assistance in the anechoic chamber and Stefan Lindgren for his technical assistance. I would also like to thank Emelie Sigrid Stiernströmer and my colleague Ariane Senecal for their support and good humor. I am grateful to Eileen Kelbach for being volunteer in my anechoic chamber recordings.

I would like to give special thanks to all the instructors at Lund University for giving precious lectures which helped me to establish my interest in linguistics.

Abstract

The study reported in the present paper aimed to explore the influence of changes in certain suprasegmental cues such as fundamental frequency and intensity on the perception of linguistic stress patterns by native speakers of American English. It attempted to determine the effect of prosodic cues on automatic word processing in the brain by comparing the mismatch negativity (MMN) component of the event-related potentials (ERP) elicited by isolated words and pseudowords. The material chosen was a pair of English words in which a change of function from noun to verb is commonly associated with a shift of stress from the first to the second syllable. Neurophysiological brain activity was recorded to series of frequent (standard) stimuli and three types of rare (deviant) stimuli differing from the standard in one of three different ways: frequency, intensity, or in both features, and the mismatch negativity (MMN) component of event-related potentials (ERP), a brain correlate of automatic preattentive auditory processing, was computed. The results of the experiment showed that in both word and pseudoword conditions, deviants elicited MMNs in a biphasic nature; one with a time course of 110-160 ms and another with a time course of 200-300 ms. These negative deflections could be interpreted to reflect the deviation of a sound from the transient auditory memory trace of the standard. However, it was unclear whether the MMNs were elicited by a change of word stress as a linguistic pattern and, consequently, lexical activation or just changes in acoustic features. Additionally, the results of the experiment showed that intensity, fundamental frequency, and combination of them contributed differentially to the prosodic information and hence, differed in their MMN amplitudes. Statistical analysis showed that the combination of the two acoustic dimensions is the most effective cue for stress perception.

Key words: language processing, brain, lexical access, word recognition, event-related potentials (ERP), mismatch negativity (MMN), stress perception, fundamental frequency, intensity

Table of Contents

List of figures

Figure number Page

List of tables

Table number Page

The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' (I found it!) but 'That's funny ...'

-Issac Asimov (1920-1992)

1. Introduction

Recent years have seen a great increase in findings about language processing in the human brain, and one level of language processing that has received increasing attention in recent psycholinguistic and neurophysiological investigations, is word recognition. Earlier research on word recognition was largely confined to the visual processing of words and, therefore, was not concerned with the segmental and prosodic cues that are crucial for auditory word recognition (Slowiaczek, 1990). It is only more recently that a growing number of studies have begun to examine the processing of segmental information in spoken words more closely. Nevertheless, the effect of prosodic information in spoken word recognition is not well understood. In spite of the fact that research investigating the use of prosodic cues in the processing of spoken words has increased in recent years, results obtained from these studies have been inconclusive (for a review see Cutler, 2005). For instance, while some investigations of the use of stress in lexical activation and recognition have failed to reveal suprasegmental effects on lexical access, some studies have shown that listeners can indeed use suprasegmental cues. Besides, even though a large number of studies confirmed the importance of suprasegmental cues in lexical access, they reported different acoustic findings, particulary with respect to the relative roles of acoustic cues. Therefore, in view of these inconclusive results from earlier studies, the present investigation attempts to obtain additional information regarding the nature of spoken word recognition in the brain in general, and examines the potential effects of suprasegmental cues on the processing of spoken words.

Several models of spoken word recognition have been developed, and most current models explain the process of spoken word recognition in terms of activation of processing units within a mental lexicon. These models consider the process of recognizing spoken words in speech as a process in which speech signals of words are matched to representations of word forms stored in the mental lexicon (Moss & Gaskell, 1999:59). In spite of the fact that there is no clear consensus about the exact structure and organization of the mental lexicon among these models, they all concur on the fact that the mental lexicon is certainly not a list of ortographic word forms. Rather, it is commonly characterized as a dictionary in which a range of information is contained in each entry (McQueen & Cutler, 1997:565). A fundamental but unsettled issue for spoken word recognition is the type of information that constrains the activation in each entry. McQueen and Cutler (1997) provide a broad overview of many of the known sources of information in this area.

The process of spoken word recognition in the mental lexicon is usually considered to consist of two major component stages (McQueen et al., 2003). The first stage is a prelexical one, where an abstract representation of the utterance is generated from the incoming information in the speech signal. The second stage is the lexical stage in which a number of word representations, which match the prelexical representation to some degree, are assumed to be activated simultaneously as candidates for an incoming word, and an appropriate one is then selected via a process of competition. In this process, incoming information is exploited as rapidly and as efficiently as possible to distinguish between competitors because the signal disappears quickly (Cutler & Otake, 2002). This incoming information potentially includes information about the basic units of prosodic or suprasegmental features[1], and these features may assist in the transformation of input to the prelexical representation (Cutler & Otake, 2002; McQueen et al., 2001; Soto-Faraco et al., 2001; Cooper et al., 2002). Accordingly, sensitivity to the parameters underlying prosodic information may be significant in establishing well-specified abstract representations of words in the mental lexicon.

Suprasegmental features of speech are defined as acoustic features that extend over more than one linguistic segment, and are usually listed as the set of features consisting of pitch, stress, and quantity (Lehiste, 1970:1). More specifically, prosodic information corresponds to the combination of perceptual dimensions: pitch, loudness, and length. The physical correlates of these perceptual parameters are fundamental frequency (*f*0), intensity, and duration (Fry, 1958). Lehiste (1979:125) states that the perception of stress appears to be based on these acoustic cues which serve as complex markers during perception. However, different acoustic parameters of suprasegmental structure may contribute differentially to the prosodic information and hence, differ in importance. Therefore, the present study addresses the processing characteristics of different acoustic parameters of suprasegmental structure, and examines the importance of these variables in the recognition of isolated spoken words and pseudowords by directly manipulating the lexical stress patterns. The purpose is to establish the differences (if any) in the relative contribution of the two acoustic dimensions: fundamental frequency and intensity in the perception of spoken words and pseudowords by native speakers of American English. The main motivation for choosing these two acoustic dimensions is the fact that the importance of vowel quality and duration as cues to English stress seem to be undisputed or less controversial in the literature. However the relative

[1] Prosodic features and suprasegmental features are used synonymously.

importance of the roles of fundamental frequency and intensity in stress perception has long been questioned (see Greenberg, 2006).

Studies that are frequently addressed to as illustrating that stress patterns affect the spoken word perception have failed to seperate the effects of suprasegmental features on their own from the correlated effects of vowel quality (for a review see McQueen & Cutler, 1997). Therefore, to investigate prosodic effects on word recognition in a lexical-stress language, it is essential to exert control over vowel quality. For instance, Fear et al. (1995) found that to English listeners, the vowel quality distinction is much more important than the purely prosodic distinction. In this respect, the most suitable material to investigate the pure prosodic effects on word recognition is minimal stress pairs, namely words with identical segments but different prosodic features. If prosodic information contributes to lexical access, in much the same way that segmental identity does, then, these minimal pairs should create distinct lexical codes, and in practice, be identifiable. Therefore, to make sure that the current research examines the effect of pure prosodic cues, a pair of English words that differ in grammatical class, where the change from a noun to a verb is linked to a shift of stress from the first to the second syllable was chosen as the experimental material.

Speech perception is a complex process and several fields of science have focused on the examination of speech perception using different methods. Stress perception and spoken-word recognition have traditionally been studied in behavioral settings but also electrophysical measures have been conducted. Since the aim of the current study is to understand the neural underpinnings of spoken-word recognition in the brain, the approach adopted is related to the auditory and neural theories and models closely connected to them. In neurophysiology, one of the most common ways of examining brain responses is measuring auditory event-related potentials (ERPs). More recently, the mismatch negativity (MMN) component of (ERPs) has been increasingly employed to examine the neural processing of speech and language (see Näätänen, 2001; Pulvermüller & Shtyrov, 2006). Speech perception requires highly organized, fast, and adaptive processes (Jacobsen et al., 2004). Since the event-related potential (ERP) technique allows one to acquire a dependent measure for assessing speech processing with millisecond accuracy and without the interference of task-related processes, it seems to be the most convenient way to investigate stress perception and spoken word recognition. Auditory sensory processes underlying sound perception can be examined using the mismatch negativity (MMN) component of auditory event-related potentials (ERPs) which is not dependent on conscious perception and response (Ceponiene et al., 2002). Therefore, in this thesis the effects of suprasegmental cues on

auditory brain responses were studied by exploiting the brain event-related potential (ERP) technique in a classic passive oddball paradigm[2], and focusing on the automatic change detection component mismatch negativity (MMN) (Näätänen, 1991; 2001).

To recapulate, the main objective of this thesis is two-fold: (1) to examine the effect of suprasegmental cues on automatic word processing in the brain, and (2) to establish the differences (if any) in the relative contribution of the two acoustic dimensions: fundamental frequency and intensity. Consequently, these objectives will be fulfilled by using event-related potential (ERP) technique, and by comparing (i) the mismatch negativity (MMN) elicited by words and pseudowords, and (ii) the MMN elicited by fundamental frequency, intensity, and their combination. The experimental material chosen consists of disyllabic words in which the location of stress on the first or second syllable led the word to be identified as either a noun or a verb, respectively. The current study involves the analysis of the output obtained from the native speakers of American English.

The general outline of the present thesis is as follows: Chapter 3 presents the background basis of this thesis and an account of related work. Chapter 3 introduces the present study, clarifying the research objectives, experimental hypotheses, and approaches to the study. Chapter 4 presents the description of the research method to be used, the description of the material and participants, the instruments to be used, the procedures, and statistical treatment utilized in the analysis. Chapter 5 contains the results of the experiment obtained from ERP data and statistical analysis. Chapter 6 presents a general discussion of findings from the perception experiment and relates the findings from the current study to existing perceptual models. In Chapters 7 and 8, limitations of the current study and recommendations for further research are discussed. Finally, Chapter 9 presents the conclusions of this thesis and explains the scientific contributions of the research, as well as its practical relevance.

2. Background

This chapter presents an analysis of the literature which addresses the key topics and concepts that are fundamental to the investigation of the current study. The literature reviewed includes a wide range of experimental documentation regarding the perception of stress, and its role in lexical access. The chapter consists of three sub-sections. The first section presents the background information of what is known about lexical stress and its lexically distictive

[2] Oddball paradigm is an experimental method in which a target stimulus is presented among more frequent standard background stimuli.

capacity. The second section reports the acoustic characteristics that guide stress perception and the relative importance of different acoustic cues by reporting the behavioral studies which examined the role of certain suprasegmental features involved in lexical access. The third section is concerned with the general characteristics of mismatch negativity (MMN), and its relevance and applications in language processing.

2.1 Lexical stress and its role in spoken word recognition

Lexical stress is defined as prosodic prominence at the word level. Terken (1991), describes prosodic prominence as the property by which linguistic units are perceived as standing out from their environment. Accordingly, word stress is prosodic prominence that specifies the relationship between the syllables of a word, so that one of these syllables is considered more prominent than the other syllables. Differences of stress are perceived by the listener as variations in a structure characterized by four perceptual dimensions: length, loudness, pitch, and quality (Fry, 1958). The physical correlates of these perceptual factors are the duration, intensity, fundamental frequency and formant structure of the speech sound waves.

Lexical stress can have an effect in speech perception due to the greater acoustic reliability of stressed syllables (McQueen & Cutler, 1997:579). Indeed, the term *lexical stress* itself implies that stress pattern can have a lexically distinctive capacity. Investigations of English vocabulary structure indicated that information obtained from lexical stress could be used in word recognition. For instance, in English, there is an association between stress pattern and grammatical function in certain classes of words; for English speakers, the word *permit*, with trochaic rhythm is a noun, and the word *permit*, with iambic rhythm is a verb (Ladefoged & Johnson, 2011:112). It has been found that listeners with no phonetic training, on hearing an isolated word of this type, can judge whether the stress is on the first or second syllable (Fry, 1958). Accordingly, lexical prosody may be of use as a possible constraint on activation in lexical access of spoken word recognition.

A number of studies have shown that word recognition is hampered when a word is pronounced with incorrect prosody (Bond & Small, 1983; Cutler & Clifton, 1984). These findings suggest that the knowledge of lexical prosody is actually stored in the mental lexicon and is used for the recognition of every spoken word. However, these studies used segmental features as well as suprasegmental features to investigate stress pattern's effect on spoken word perception, and as a consequence, failed to distinguish the effects of suprasegmental features from the associative effects of vowel quality. For instance, Bond and Small (1983)

found that word recognition in shadowing[3] was fulfilled in spite of mis-stressing as long as the mis-stressing did not result in a change of vowel quality. In a related study, Cutler and Clifton (1984) showed that when two-syllable words were mispronounced by shifting lexical stress pattern, subjects' reaction times to correctly stressed words were significantly faster than reaction times to incorrectly stressed words. Correspondingly, they suggested that word recognition was hampered when a word was misstressed. However, Cutler and Clifton also indicated that shifting stress without altering vowel quality had a much smaller adverse effect on recognition of stress shifts which changed full vowels to reduced or vice versa. All these results indicate that segmental information outweighs suprasegmental information in lexical activation in English. This wide range of evidence could mean that English listeners make very little use of suprasegmental stress cues in lexical access.

To investigate whether lexical prosody has a role in lexical activation or not, Cutler (1986) conducted an experiment using a cross-modal priming task with English participants. She examined English heteronyms, such as FOREbear (meaning ancestor) vs. forBEAR (meaning withhold), in which stress is contrastive but there are no segmental differences (there is no vowel reduction in either word). Results from this priming study indicated that English speakers could not utilize prosodic stress information alone to achieve lexical recognition. As a reason for this finding, Cutler (1986) pointed out the fact that English has only a very small number of minimal stress pairs in which vowel reduction is not employed.

In a related study, Cutler and van Donselaar (2001) conducted the same experiment for Dutch, which has an equally small number of minimal stress pairs, and found the opposite result. Accordingly, they suggested that this opposite result may imply that Dutch speakers used prosodic information in identification of lexical items. In light of these two different findings, the authors argued that vowel quality outweighs other prosodic information in recognition of stress contrasts in English but not in Dutch because Dutch and English differ in the amount of vowel reduction (Dutch has much less vowel reduction than English). Correspondingly, Cooper et al., (2002) suggested that English simply does not provide its listeners with as much possibility for suprasegmental processing in word recognition as some other languages do. English listeners will rarely experience a segmentally ambiguous but suprasegmentally disambiguated fragment of speech. In light of this, it can be stated that where segmental information distinguishes words more rapidly than suprasegmental

[3] Speech shadowing is an experimental technique used in psycholinguistics

information, there may be little encouragement for listeners to attend to the suprasegmental features.

2.2 The Perception of Stress

There is a large body of literature dealing with the acoustic correlates of stress. For many years, researchers have attempted to identify the acoustic dimensions being perceived when listening to stress. English has been the focus of many investigations, especially since the 1950s, and a number of studies have examined the acoustic correlates of lexical stress in American English. Some of these studies have become pivotal in the literature on the perception of stress. Most of these studies focused on lexical stress in English disyllabic words in which the location of stress on the first or second syllable led the word to be identified as either a noun or a verb, respectively. Results of these studies consistently indicate that the acoustic correlates of average fundamental frequency (f0), intensity, and duration are associated with the perception of English lexical stress: Stressed syllables have higher f0, greater intensity, and longer duration than unstressed syllables (Fry, 1955; 1958). The listeners, in normal conditions, have a number of cues that they can use as the basis of any single judgment and these cues are provided by variations in any or all of the perceptual dimensions. However, the listeners may, for a specific judgment, be more dependent on one than on another. The relative importance of intensity, fundamental frequency, and duration in the perception of stress has been studied experimentally in English. In the following section, we will look at some of these studies.

2.2.1 Behavioral studies of English stress perception

Fry (1955, 1958) studied English stress extensively and explored the influence of certain physical cues on the perception of linguistic stress patterns. He ran experiments in order to measure the effect of changes in three physical dimensions; duration, intensity, and fundamental frequency on stress judgements of English listeners. In his classic study from 1955, Fry conducted a series of experiments with English disyllabic noun-verb word pairs in which a change of function from noun to verb is commonly associated with a shift of stress from the first to the second syllable (as in *object, permit, digest*). These words were recorded, resynthesized and then played to native speakers of English to determine the relative effectiveness of duration and intensity as cues for stress judgements. The results of these experiments were much what one would expect. They indicated that duration and intensity ratios were both cues for judgements of stress. The vowel segments showed the major

differences in duration and intensity with a shift of stress. When the vowels were long and of high intensity, they were judged as stressed: when they were short and of low intensity, they were considered unstressed. The most interesting aspect of the results was shown when the effects of duration and intensity were studied on their own. The results showed that the duration ratio had a stronger influence on judgements of stress than the intensity ratio had.

In his 1958 study, Fry tested all three dimensions: duration, intensity and fundamental frequency, in a series of experiments, as cues for the perception of stress. In order to experiment with judgements of stress, Fry used the same word pairs as in his 1955 study, but used a slightly different method. In the first experiment, he combined variations of duration and intensity. The second experiment combined duration changes with step changes of fundamental frequency (*f0*). The third experiment included variations in fundamental frequency within one syllable and contained a range of patterns which imposed sentence intonation on test items. The results of these experiments confirmed the previous findings about English stress being marked by first duration then intensity. Changes in *f0* were also found to be more crucial than the magnitude of the changes themselves. Fry indicated that change in fundamental frequency differs from change of duration and intensity in that it tends to produce an 'all-or-none' effect, that is to say the magnitude of the frequency change seems te be relatively unimportant while the fact that a frequency change has taken place is all-important. According to Fry, the hierarchy of stress correlates is therefore fundamental frequency, duration, and intensity, with fundamental frequency being the most important in determining stress.

Morton and Jassem (1965) also attempted to identify the acoustic correlates of stress in English and got similar results with the previous experiments. In their study, the parameters of fundamental frequency, intensity, and duration were varied systematically. The stimuli were presented to English speakers. The results showed again that fundamental frequency was by far the most important physical correlate of perceived stress. Variations in fundamental frequency produced greater effects than variations in either intensity or duration. The results of Morton and Jasses confirmed the 'all-or-none' effect of fundamental frequency changes already observed by Fry (1958).

All the studies reviewed up to now have been devoted to establishing the perceptual significance of various suprasegmental cues incorporated into listening tests in which the task has been to identify the stressed syllable. In summary, it appears that when vowel quality is controlled, fundamental frequency provides the primary cue for the presence of stress. Native English speakers mostly rely on the fundamental frequency cue to detect stressed segments in

the speech signal. Duration also appears to play a significant role and, to a lesser extent, intensity.

2.3 Background for Mismatch negativity (MMN)

The human brain encodes physical features of sounds in neural representations stored in auditory sensory memory (Nelson, 1998:71). Auditory sensory memory (ASM) is a critical first stage in auditory perception that permits the listener to integrate incoming acoustic information with stored representations of preceding auditory events. It is considered to be the earliest memory store where pre-conscious sound feature analysis, integration, and discrimination occur (Ceponiene et al., 2002). This memory module provides a database for information processing mechanisms. The representations in this database last for several seconds and are thought to involve a complex and widely distributed neural network (Alain, Woods & Knight, 1998). Recently, this memory system has been extensively studied by using the mismatch negativity (MMN) component of event-related potentials (ERPs).

The mismatch negativity (MMN) is an electrophysiological measure that reflects differences in sound discrimination sensitivity, and indexes auditory change detection. The MMN was discovered by Näätänen et al. (1978). The MMN is elicited by infrequent deviant stimuli occasionally replacing frequently occurring, 'standard' stimuli (Näätänen et al., 1978; for a review see Näätänen, 2001) and generated by a brain mechanism that compares each new auditory input with a neural trace of the repetitive stimulation held in the memory. The MMN displays a frontocentral scalp distribution and a latency between 100-250 ms after deviance onset. MMN can be elicited in the absence of the subject's attention to the auditory input (Tiitinen et al., 1994) and therefore, signals the brain's automatic reaction to any change in the auditory sensory input (Shtyrov et al., 2009). The MMN is thought to reflect a fairly automatic process that compares incoming stimuli to a sensory memory trace of preceding stimuli. The fact that the MMN response is larger in amplitude for familiar than unfamiliar speech sounds indexes the participation of long-term memory in MMN elicitation (Näätänen, 2001). In addition to speech sounds and their combinations, the MMN has been employed to investigate representations for words in the brain (Pulvermüller et al., 2001). They found that an infrequent syllable completing a word in a sequence of syllables elicited an MMN that was larger in amplitude in comparison with a syllable completing a pseudoword. The authors argued that the MMN enhancement could be explained by the activation of memory traces for words.

The MMN is usually calculated as ERPs evoked by a standard stimulus are subtracted from ERPs evoked by the presentation of a deviant stimulus (Näätänen, 1992) (Fig. 1)[4].

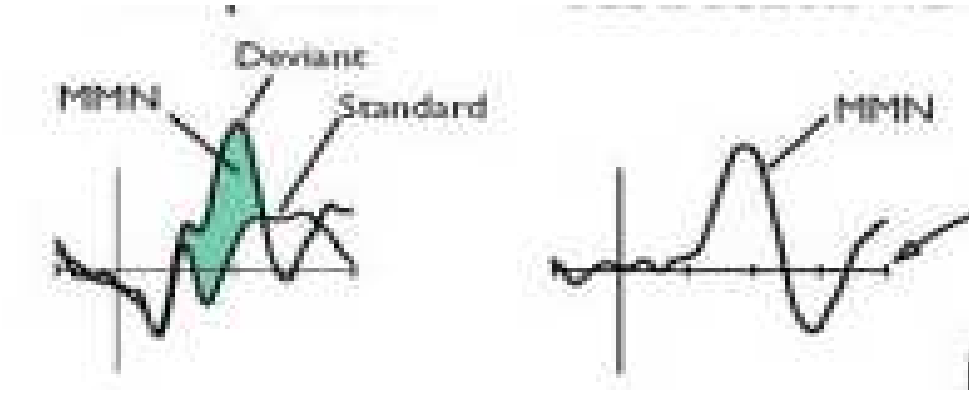


Figure 1 The MMN wave is a subtraction of the ERP to the standard stimulus from the ERP to the deviant stimulus. Note that in this figure, the polarity is inverted (negative up).

In contrast to other ERP components, the MMN provides a physical measure of the actual sensory information processed by the brain (Näätänen, 1992:136). Changes in repetitive physical features of the acoustic input elicit the mismatch negativity. The elicitation of the MMN can be explained by the 'memory-trace hypothesis' according to which our brain encodes physical features of the acoustic input into short-lived neural traces (i.e. sensory memory), establishes representations of repetitive features in the acoustic input as a neural model, and compares each input with this model (Schröger, 1996). If a difference is detected by this memory comparison process, the MMN is elicited. Hence, MMN can be a practical measure to study auditory sensory memory and comparison processes in the memory. Additionally, since the MMN is also elicited when subjects do not actively listen to the auditory input, it is especially appropriate for studying attention-independent processing involved in auditory sensory memory and change detection.

Mismatch negativity research has indicated that the brain is capable of extracting complex and abstract rules about regularities that enable the auditory system to expect the most likely attributes of the next sound (Näätänen et al., 2001). The simplest pattern of such sequences consists of a repeating identical sound with MMN elicited to rare deviants in which the sound is changed in some physical features. Given input of this kind, the brain builds a model of the attributes that define the regularity, and this model functions as an asumption of the most likely properties of the succeeding sound (Winkler et al., 1996). Any violation of a

[4] For a review of the measurement and interpretation of mismatch negativity, see Schröger, 1998.

sound pattern elicits the MMN. This inferential process is automatic, occurring even during sleep (Atienza et al., 1997) and is generally investigated while participants ignore sounds and pay their attention to another task (Näätänen et al., 1993).

As stated above, mismatch negativity (MMN), considered as an automatic index of experience-dependent auditory memory traces, has been employed for investigating the neural processing of speech and language. There are a number of studies which have indicated word specific activations in the absence of focused attention on the stimuli, thus signifying automatic brain access to items stored in the mental lexicon. These studies found that the MMN elicited by spoken words is larger than that elicited by comparable pseudowords, suggesting that the MMN amplitude can indicate the presence of memory traces or cell assemblies for words, representing spoken words in the human brain (Shtyrov & Pulvermüller, 2002; Pulvermüller et al., 2004).

2.3.1 The MMN for basic stimulus features

As already mentioned, any discriminable auditory change elicits an MMN (for a review see Näätänen & Winkler, 1999). In order to assign the MMN as a probe into auditory sensory memory and automatic change detection, its dependence on specific physical stimulus attiributes should be explained. Correspondingly, in this section, several studies where the MMN for changes in basic stimulus features (frequency, intensity, duration) are addressed. In these studies it was confirmed that the MMN can be elicited by both increments and decrements in basic stimulus features (for a thorough review, see Näätänen, 1992).

The MMN reflects processing of minimal acoustic differences (Sams et al., 1985), and it can be obtained in response to changes in various stimulus parameters such as fundamental frequency, intensity, location and duration. Therefore it may index a neuronal representation of the discrimination of several auditory stimulus properties (Näätänen et al., 1978). Moreover, it can be obtained when the acoustic differences between the standard stimuli and the deviant stimuli are small enough to be near the psychopsychological threshold. For instance, responses have been obtained when the stimulus difference is as small as 8 Hz (0.8 % change) and 6 dB (Sams et al., 1985; Näätänen et al., 1987).

2.3.2 MMN for word stress

The MMN to word stress has been investigated by a number of studies. In a recent study, Ylinen et al. (2009) examined the effect of prosodic familiarity on automatic word processing in the brain by comparing the mismatch negativity (MMN) elicited by words and

pseudowords with familiar and unfamiliar stress patterns. They found that the MMN was elicited by a change from unfamiliar to familiar words and a change from a familiar to an unfamiliar word-stress pattern. When familiar words were accompanied by an unfamiliar stress pattern, the MMN response was considerably hindered in comparison with the familiar words with a familiar stress pattern. They concluded that an unfamiliar prosodic pattern increased the computational needs in word recognition but did not hinder recognition.

In another study, Weber et al. (2004) studied German infants' sensitivity to stress, by examining the mismatch negativity (MMN) elicited by changes in trochaic and iambic patterns. The result showed that with respect to the stress patterns investigated, MMR are only observable for trochaic items by the age of 5 months. The authors based this result on the fact that the trochaic stress pattern is more frequent in the target language and therefore, might be detected more easily by infants at this age.

Adults seem to process stress pattern violations automatically, and most likely depend on matching processes with regard to abstract representations of the regular stress pattern in the mental lexicon. This hypothesis was supported in a recent MMN study by Honbolygo et al. (2004). They examined the MMN activations to mis-stressing in bisyllabic words, and found that stress violations of bisyllabic words lead to the occurrence of two different MMNs, one to the stress withdrawal on the first syllable and one to the additional stress on the last syllable.

A number of MMN studies on stress (Honbolygo et al., 2004; Weber et al., 2004; Fowler et al., 1986) raised at least two important questions about language development regarding language specific rhythmical sensitivity. First, what is the role of acoustic saliency in processing stress patterns of spoken utterances and how might this contribute to the phonological development? Second, what kind of interaction is occuring during development that starts as a general sensitivity to all possible speech rhythms and gets shaped by language experience? Consequently, investigations indicated that language-specific shaping is mediated by the rhythm and stress-pattern representations stored in the phonological lexicon, together with the phonemic and phonotactic attributes of words in a specific language.

3. The Present Study

3.1 Research objectives and approaches to the study

As mentioned in the introduction, the present research is oriented towards contributing to knowledge regarding potential effects of lexical stress on the processing of spoken words. The

main objective of the research is to determine the role of acoustic saliency (consisting of several acoustic features) in processing stress patterns of spoken utterances and how much this system contributes to lexical activation. The research objective is achieved by undertaking an electrophysiological research approach, through an ERP study. The aim is to find the primary perceptual cues to word stress in English and the acoustic parameters that change in response to change in stress placement at the word level. To our knowledge, no studies have examined the relative effectiveness of prosodic cues in stress contrastive verb-noun pairs in English using a brain imaging technique.

More specifically, the current study addresses the following questions: (i) does prior knowledge of placement of lexical stress speed up word recognition, and (ii) how do native listeners weight pitch and intensity in relation to each other in stress perception? To answer these questions, we recorded the ERPs for a pair of English disyllabic words, in which the location of stress on the first or second syllable led the word being identified as either a noun or a verb, and a pair of pseudowords that have the same stress features as the real words. The reason for choosing stress-contrastive verb-noun pairs is that the shift from the one syllable to another syllable will still result in a meaningful word and this makes it possible to look at the lexical activations in the brain as a result of a change in stress placement.

MMNs were recorded in relation to the second syllable of words and pseudowords in an oddball paradigm, by presenting three different deviants interspersed among standard stimuli that occur on most trials. One deviant differed from the standard on pitch features, a second deviant differed from the standard on intenstiy features, and a third deviant differed from the standard on both features.

3.2 Experimental Hypotheses

The present study attempts to test four hypotheses about the role of acoustic cues in spoken word recognition and their relative importance.

3.2.1 Deviations from acoustic regularities:

The first hypothesis is that MMN elicitation should be realized as early as ~110 ms since the MMN is elicited in the auditory cortex when incoming sounds are detected as deviating from a neural representation of acoustic regularities. Since the MMN implies the existence of an auditory sensory memory that stores a neural representation of the standard against which any incoming auditory input is compared, deviations from standards should elicit MMNs predictably.

3.2.2 Early recognition of words:

The second hypothesis is related to the role of prior knowledge of lexical stress placement and acoustic realization for word recognition speed. If listeners can use anticipatory information about stress pattern, they should recognize words more quickly than pseudowords. Accordingly, MMNs elicited for real words should be higher in amplitude and earlier in latency, suggesting the early and larger activation of lexical memory traces. This hypothesis is based on the assumption that the preexisting memory networks of words which have strong internal connections guarantee rapid activation. Since the pseudowords do not have such preexisting memory networks, no early acivation is expected for them.

3.3.3 Distinct lexical codes in neural representations:

If word stress information participates in lexical access, then, the minimal pairs should generate distinct lexical codes, and automatic brain access to these items in the mental lexicon should be realized. Listener's noun vs. verb judgments should indicate to what extent the stress pattern is responsible for lexical activation. For instance, a stress change from the second to the first syllable should elicit a higher MMN in real words since in both cases, the result will be a meaningful word, i.e. hearing a noun rather than a verb.

3.3.4 Relative importance of different acoustic cues

The fourth hypothesis is that different acoustic parameters of suprasegmental structure may contribute differentially to the stress-related information and hence, differ in their MMN amplitudes. Listener's noun vs. verb judgments indicate to what extent the variable under consideration is responsible for stress recognition. The most important cue should elicit the higher MMN and in the light of earlier research, fundamental frequency and the combination of fundamental frequency and intensity should elicit a higher MMN.

4. Method

4.1 Participants

The experimental group consisted of twelve native speakers of American English (8 females, 4 males, 3 left-handed). Their age ranged from between 20 to 32 years (M=26). All of them were born and grew up in the United States. All were studying at Lund University and had at least 15 years of formal education. All participants reported normal development and hearing, and none of them had a history of neurological diseases or language-related disorders. Prior to

the experiment, each participant signed an informed consent form explaining the study and stating that the participant was free to leave the study at any time, and for any reason. All the participants were included in the entire analysis[5]. All participants also answered a questionnaire regarding their background upon arriving at the laboratory. The participants were recruited from the Lund University community via a small poster advertisement circulated at the university campus and a message posted to the social network of the International Student Desk.

4.2 Materials

4.2.1 Stimulus recording

Six pairs of English disyllabic words, in which the location of stress on the first or second syllable led the word to be identified as either a noun or a verb, were recorded as alternatives for the material that would be utilized in the present experiment. Each pair was produced by a female native speaker of American English (originally from Ohio, 26 years old), and none of the pairs involved any differences in vowel quality. Each pair was elicited under three conditions: (i) in isolation, (ii) in the semantically neutral frame sentence, and (iii) in associated context sentences created specifically for each word (for the whole recording script, see Appendix 2).

The speaker was recorded in the anechoic chamber in the Humanities Lab at Lund University. The microphone was placed approximately 20 cm from the speaker's lips at an angle of 45º (horizontal) during recording. The speaker was instructed to speak at a natural rate and loudness level. The speech tokens were sampled at a rate of 44.1 kHz with a quantization of 16 bits and low-pass filtered at 22.05 kHz. Each token was then normalized and saved as an individual sound file.

After the recordings, a pair of utterances of *UPset* (noun) and *upSET* (verb) was selected to be an exemplar in the stress perception experiment. The specific token was chosen on account of the fact that it was acoustically clear enough to display identifiable pitch and intensity contours, which could guarantee the technical manipulation for the perception experiment. After settling upon the material, a pseudoword pair, *UKfet* (noun) and *ukFET* (verb), mimicking the acoustics of the real word pair was created by slightly modifying the

[5] During the EEG recordings, two of the participants were sleepy, but they are included in the analysis since the inferential process of MMN is automatic, occurring even during sleep (Atienza et al., 1997)

selected word pair, and recorded in anechoic chamber. The motivation for having pseudowords was that a pseudoword contrast would make it possible to see whether the effects obtained were associated with some kind of lexical activation, or just because of the changing acoustic features.

4.2.2 Stimuli manipulation

For manipulation, the recordings were transferred to the Praat speech analysis and synthesis software programme 4.5.12 (Boersma & Weenink, 1992-2009)[6]. Using Praat, the following acoustic parameters were measured for each token: average intensity (in dB) and average fundamental frequency (f0 in Hz). The parameters related to intensity were measured within a syllable while f0 parameters were measured over the vowel.

The first step was to create a single version of the *upset* token such that it would be perceived as ambiguous with respect to stress placement. This was done by concatenating the *up-* of verb *upset* and *-set* of the noun *upset*. To make sure that the token was totally neutral, pitch and intensity for both syllables were manipulated so as to have the same values as the averaged measures derived from the original stressed and unstressed syllables of *upset* pair (Tables 1 and 2).

Table 1

Values of acoustic parameters of the syllable *up-* in *upset* in stressed and unstressed conditions produced by an American native speaker. The average of the acoustic parameters is derived from the (stressed+unstressed) values/2

Up	f0 (Hz)	Int (dB)
Stressed (noun)	242 Hz	79 dB
Unstressed (verb)	189 Hz	73 dB
Average	215 Hz	76 dB

Table 2

Values of acoustic parameters of the syllable *-set* in *upset* in stressed and unstressed conditions produced by an American native speaker. The average of acoustic parameters is derived from the (stressed+unstressed) values/2

Set	f0 (Hz)	Int (dB)

[6] www.praat.org

Stressed (verb)	206 Hz	77 dB
Unstressed (noun)	172 Hz	73 dB
Average	189 Hz	75 dB

Using Praat, the revised neutral token was manipulated to create the standards and deviants (See Appendix A for a thorough description). To create the standard upset, leaning towards a verb, the second syllable of the neutral token *–set* was resynthesized, and the f0 contour (202 Hz) was enhanced 10 Hz[7] and set to 212 Hz., and the Intensity was enhanced 1 dB[8]. This should be an audible difference considering that the just-discriminable difference is approximately 1 dB (Lehiste, p.121). To create the deviants, necessary changes were made to the second syllable by lowering the values[9] and the first syllable *–up* was held constant with the values of the baseline neutral token. To create the pitch deviant, the second syllable was resynthesized by lowering the pitch value from 212 Hz to 180 Hz., based on the originally recorded unstressed values The intensity was held constant at the value of the standard. To create the intensity deviant, the intensity of the second syllable of the standard was manipulated to have the same averaged value derived from original unstressed syllables of the *upset*-pair, 73 dB. To create the third verb deviant, the first two deviants were combined. It is possible that the relatively long duration of the second syllable *-set* may influence listeners such that they will perceive the second syllable stressed. In order to offset this, the second syllable '-set' was shortened by 50 ms, resulting in a total length of 425 ms. The same steps as previously used were applied for the pseudoword pair. In the following table the manipulation results can be seen.

[7] Since the perception of pitch is not uniform with regard to frequency change and is also dependent upon the duration and intensity of the signal, determining the threshold of perceptible change of *f*0 is more complex (Lehiste, 1970:64). Thus the present author chose the lowering value for pitch "10 Hz" arbitrarily.

[8] The precise measure for computing intensity has been debated. Fry (1955, 1958) and Beckman (1986) identified average intensity over the syllable as a possible acoustic correlate of stress differences, while others (Sluijter and van Heuven, 1996) have argued that the frequency spectrum of a given vowel is a more appropriate measure. In the current study only average intensity was used.

[9] Since it is likely that MMN responses will be derived by stronger acoustc deviance, to avoid that an increase in auditory features giving an intrinsic increase in the MMN, the deviants were created by lowering the intensity/ pitch of the second syllable

Table 3 Complete stimulus design

	BLOCK A (word) [upset]			**BLOCK B (pseudoword) [ukfet]**	
Standard (verb)	up	–	set	uk	–	fet
Pitch	202 Hz		212 Hz	202 Hz		212 Hz
Intensity	75 dB		76 dB	75 dB		76 dB
Pitch deviant (noun)	up	–	set	uk	–	fet
Pitch	202 Hz		180 Hz	202 Hz		180 Hz
Intensity	75 dB		76 dB	75 dB		76 dB
Intensity deviant (noun)	up	–	set	uk	–	fet
Pitch	202 Hz		212 Hz	202 Hz		212 Hz
Intensity	75 dB		70 dB	75 dB		70 dB
Pitch+Intensity deviant (noun)	up	–	set	uk	–	fet
Pitch	202 Hz		180 Hz	202 Hz		180 Hz
Intensity	75 dB		70 dB	75 dB		70 dB

4.3 Procedure

The experiment was run using E-Prime version 2.0.1.06 (Psychology Software). Following the electrode application, participants were seated in front of a computer screen and the stimuli were presented auditorily via two loudspeakers that played sound at a comfortable listening level of 60-65 dB sound-pressure level (SPL) in a quiet lab. Prior to the experiment, the experimenter informed the participants of the stress shift rule in English disyllabic words, and they clearly claimed their understanding of the association of the stress location on the first and the second syllable of English disyllabic words to the noun-verb category.

Stimuli were presented in an auditory passive oddball paradigm which involves presenting a regular train of frequent (standard) and rare (deviant) stimuli differing in some

discriminable way. Since, the study was designed to investigate possible neurophysiological differences between individual spoken words in the absence of focused attention to these stimuli, participants were asked not to attend to stimuli, which were delivered through loudspeakers. To distract them from the language input, they were shown during ERP data collection a silent documentary[10] that did not include subtitles. To guarantee the withdrawal of attention from the lexical material, the participants were told beforehand that after the experiment they would be asked to fill out a small questionnaire, in which they would have to answer a number of questions about the documentary. In order to minimise the eye movements, the documentary video shown only filled approximately a quarter of the screen. Instructions were given both orally and in written form, appearing both in paper form and on the computer screen before the initiation of each study procedure.

Stimuli were presented in a random order (standard: P=8/10, deviant: P=2/10) and the stimulus onset asynchrony (SOA) was set 1000ms. The offset-to-onset interstimulus interval (ISI) was 575 ms and stimulus duration was 425 ms. Figure 2 below illustrates the imagined presentation of the stimuli in the oddball paradigm and shows ISI, SOA, and stimulus duration.

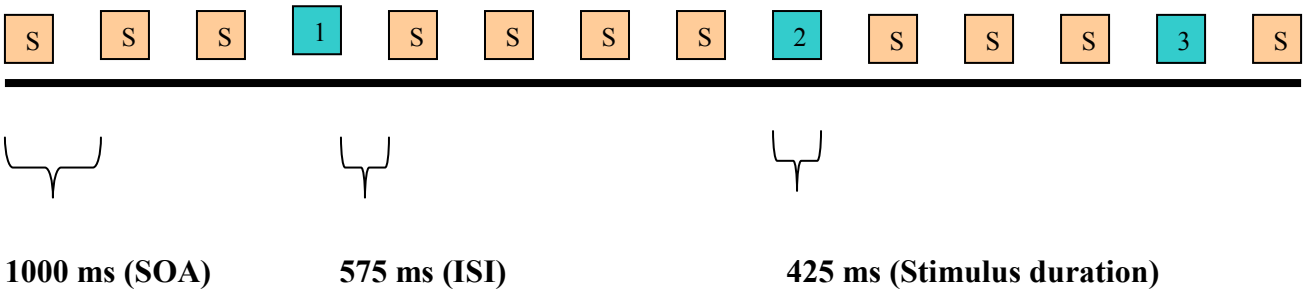


Figure 2 Presentation of stimuli in a classic oddball paradigm. Green boxes are deviants.

The experiment consisted of two blocks, each block consisting of 720 standard and 180 deviant (60 for each deviant) trials. In Block I (word block), lowered pitch, lowered intensity and lowered pitch+intensity were presented as infrequent deviant stimuli, against the background of the frequent standard stimulus *upset*. In Block II (pseudoword block), lowered pitch, lowered intensity and lowered pitch+intensity were the deviants and the pseudoword *ukfet* the frequent standard stimulus. Each block of the experiment took 15 (min) of running time (900x1000=900,000 ms/ 15 min) and there was a break between the blocks. The order of the two blocks was counterbalanced across the participants. The experimental procedure took

[10] 'Planet Earth' – Nature documentary produced by the BBC Natural History Unit (2006)

approximately 2,0 h, including the application and removal of electrodes. All participants were instructed to remain as still and relaxed as possible during the recording blocks.

4.4 EEG recordings

The acquisition of EEG signals was performed by means of NeuroScan Acquie Software, using a SynAmps 2 amplifier, and a 32-channel Easy Cap (Figure 3). Recording electrodes were mounted in an electrode cap according to the 10-20 placement system at the following positions: Fp1, Fp2, F7, F3, Fz, F4, F8, FC3, FCz, FC4, FT7, FT8, T7, T8, TP7, TP8, C3, Cz, C4, TP7, TP8, T7, T8, CP3, CP4, P7, P3, P4, P8, O1, Oz, O2, M1, M2. The reference electrode was at point Pz, and the ground electrode was placed between Fz and Fpz, on the midline. Eye movements were monitored with two horizontal eye-electrodes (HEOGL and HEOGR) and vertical eye-electrodes at the left eye (VEOGU and VEOGL). The contact impedance was kept below 5 kΩ at each electrode site. The recording bandwidth was 0.1-70 Hz and the sampling rate was 500 Hz. The channels were referenced to the cap-mounted online reference electrode while recording, but were re-referenced to the average of left and right mastoids offline.

Figure 3 Synamps 2 amplifier and NeuroScan Easycap

4.5 ERP data analysis

Offline analysis of the data was done with Neuroscan Edit Software comprising the following steps. After applying a band-pass filter (0.5-30 Hz, 24 dB/oct), the raw EEG data were segmented into epochs of 900 ms, time-locked to the onset of the critical syllables (the second) (100 ms before onset to 800 ms after onset) and stimulus-triggered event-related potentials were calculated for each participant, electrode, and stimulus. The segmentation was performed separately in each block. The critical syllable started at 135 ms in word stimuli and at 126 ms in pseudoword stimuli. In the word stimuli, stimulus-triggered event-related

potentials (ERPs), starting at 35 ms and lasting until 935 ms, in pseudoword block, starting at 26 ms and lasting until 926 ms were calculated. Subsequently, the channels were re-referenced to the averaged mastoids. Next, the segmentation data were baseline corrected from 100 ms to the onset of the critical syllable. Artifacts from eye movements were searched for on VEOGU and ocular artefact reduction was performed. Artifact rejection was set to omit activity exceeding ±100μV at any channel. Finally the ERPs were averaged separately for each stimulus type and condition, and grand averages were computed independently for each of the stimulus types. Grand average ERP waves were obtained by averaging the individual waveforms of 12 participants. The present study used a criterion of 15 minimum sweeps value[11] (accepted trials as minimum) in order to be included before grand averaging. The mean numbers of accepted trials are given in Table 4.

Table 4: Numbers of accepted trials for two blocks

Statistics of accepted trials in word block					
	N	Minimum	Maximum	Mean	Std. Deviation
standard	12	234	570	427,08	127,000
pitch deviant	12	18	50	37,08	11,790
intensity deviant	12	19	50	38,58	11,509
pitch+intensity deviant	12	21	50	35,58	10,440

Statistics of accepted trials in pseudoword block					
	N	Minimum	Maximum	Mean	Std. Deviation
standard	12	215	630	422,42	151,231
pitch deviant	12	18	54	35,42	13,090
intensity deviant	12	15	52	36,58	12,965
pitch+intensity deviant	12	19	57	34,83	13,347

As shown in Table 4, accepted trials for both word block and pseudoword block had approximately the same numerical values for all of the four conditions: standard, pitch deviant, intensity deviant, and pitch+intensity deviant. The ERP waveforms were quantified by computing the mean amplitudes of averages. Average values of the MMN curves in predefined intervals were calculated for each deviant, participant, and recording site and submitted to statistical analysis. The choice of latency windows was based on the grand

[11] Even though 15 as a minimum sweep value is very small, it is considered enough for the current study.

averages. The measurement windows were determined by visual inspection of grand average waveforms and previous findings. The MMN amplitude was measured from the different waves, obtained by substracting the average standard stimulus response from the average deviant stimulus response separately for each subject and electrode, as the mean amplitude.

4.6 Statistical analysis

Statistical analysis was performed in SPSS (IBM® SPSS® Statistics Version 19). The experimental hypotheses formulated in section 2.3 were tested independently. For the analysis, frontal (F), fronto-central (FC), and central electrodes (C) were used because the MMN response is usually reported to be most prominent at the fronto-central sites (Näätänen, 1999). Accordingly, statistical analysis of MMN amplitudes was restricted to a grid of 9 frontocentral electrodes, corresponding to the positions F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4 (Figure 4).

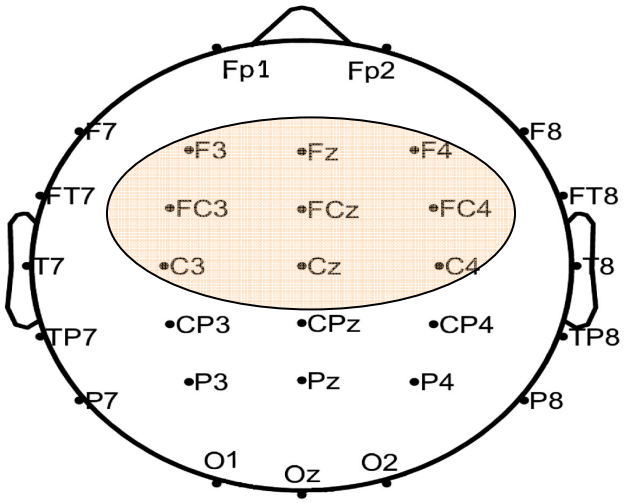


Figure 4. Channel location and the nine channels chosen to further statistical analyses: F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4.

Initially, mean amplitudes were assessed via three-way repeated measures ANOVAs on each electrode site, in the relevant time windows. The factors were *lexicality* (two levels: word and pseudoword), *prosody* (four levels: standard, pitch deviant, intensity deviant, pitch+intensity deviant), and *electrode site* (9 levels). In case of significant Prosody x Lexicality, follow-up one-way ANOVAs were performed for both words and pseudowords separately in order to investigate processing differences between conditions. Greenhouse-Geisser corrected degrees of freedom are reported when the assumption of sphericity was not met for the within-subject factors. Otherwise, the degrees of freedom are reported with sphericity assumed. In addition

effect sizes are reported (partial eta-squared: η^2), using the commonly used guidelines (.01=small, .06=moderate, .14=large effect).

5. Results

5.1 ERP data

A preliminary visual inspection suggested two different MMNs both in the word block and the pseudoword block; one with a time course of 110-160 ms and another with a time course of 200-300 ms after the onset of the second syllable. Figure 5 shows MMNs recorded at the frontal (Fz), fronto-central (FCz) and central (Cz) electrodes for the second, critical syllables presented in word and pseudoword contexts.

Word Block **Pseudoword Block**

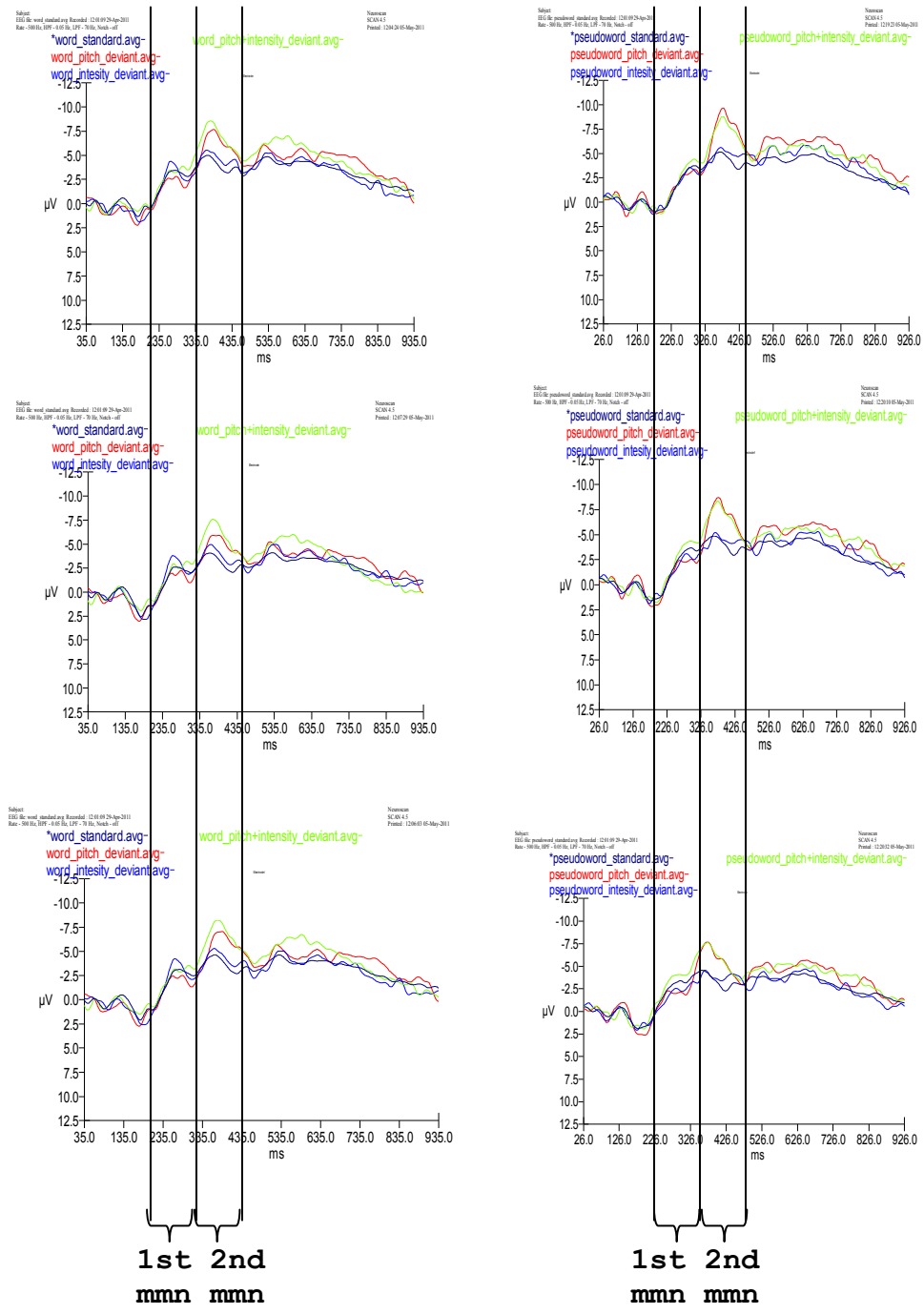


Figure 5 MMNs elicited by the standard and deviant stimuli in the two experimental blocks at channels Fz, FCz, and Cz respectively.

In the word condition, the critical syllable started at ~135 ms, and in pseudowords at ~126 ms. In the word condition, the earlier and smaller wave started at ~245 ms (110 ms after the onset of the second syllable) and continued for about ~50 ms till ~295 ms. after the onset of the critical syllable. The later and larger MMN wave started at ~335 ms (200 ms after the onset of the second syllable) and continued for about ~100 ms till ~435 ms. In the pseudoword

condition, the earlier and smaller wave had a time course from 236 ms (110 ms after the onset of the second syllable) to 286 ms (50 ms after the start of the effect). The later and larger MMN wave had a time course from 326 ms (200 ms after the onset of the second syllable) to 426 ms (100 ms after the start of the effect) (Figure 6).

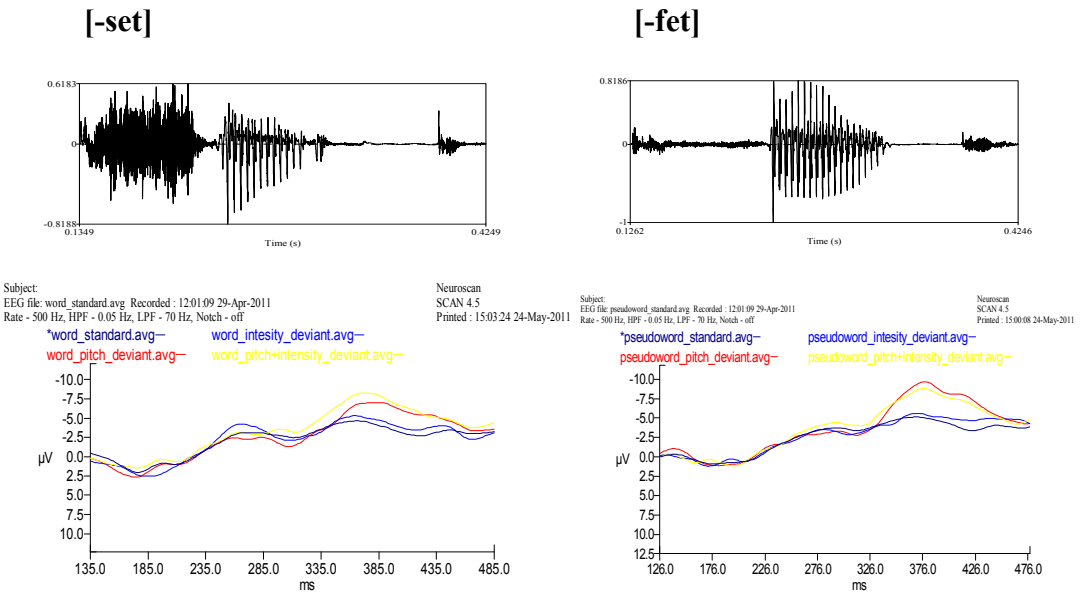


Figure 6 MMNs to the second syllables at electrode site FZc

5.2 Statistical Analysis

5.2.1 110-160 msec Latency Region

A repeated measures of ANOVA was conducted to compare MMNs elicited for the standards and deviants at the 110-160 ms latency region after the onset of the critical syllable.

Table 5 ANOVA and Bonferroni corrected t-test table for 110-160 ms after the beginning of the critical syllable

Factor	F	p	partial eta squared
Lexicality	F(1,11)=.392	P=.544	.034
Prosody	F(3,33)=.646	p=.591	.055
Electrode site	F(8,88)=12.115	p=.000*	.524
Lexicality x Prosody	F(3,33)=.502	p=.684	.044
Lexicality x Electrode site	F(8,88)=1.001	p=.390	.083
Prosody x Electrode site	F(24,264)=1.931	p=.007*	.149

The repeated measures of ANOVA with a Greenhouse-Geisser correction determined that mean MMN amplitudes did not show a main effect for Lexicality (F(1,11)=.392, p=.544) and Prosody (F(3,33)=.646, p=.591). However there was a Prosody x Electrode site interaction (F(24,264)=1.931, p=.007). Since pre-defined electrode sites were included in the statistical analysis, follow-up ANOVAs for electrode site interactions were not conducted.

5.2.2 200-300 msec Latency Region

Another repeated measures ANOVA was carried out to compare MMNs elicited for the standards and deviants at 200-300 ms latency region after the onset of the critical syllable.

Table 6

ANOVA and Bonferroni corrected t-test table for 200-300 ms after the beginning of the critical syllable

Factor	F	p	partial eta squared
Lexicality	F(1,11)=1.178	p=.301	.097
Prosody	F(3,33)=16.920	p=.000*	.606
Electrode site	F(8,88)=22.108	p=.000*	.668
Lexicality x Prosody	F(3,33)=.654	p=.586	.056
Lexicality x Electrode site	F(8,88)=.173	p=.994	.015
Prosody x Electrode site	F(24,264)=3.098	p=.000*	.220

There was not a main effect for Lexicality (F(1, 11) = 1.178, p = .301). However, the output showed that there was a main effect for Prosody (F(3,33)=16.920, p=.000) and a Prosody x Electrode interaction (F(24,264)=3.098, p=.000), but no significant Prosody x Lexicality interaction (F(3,33)=.654, p=.586). In the follow-up ANOVAs, the effect of prosody was examined for words and pseudowords separately (Figure 7 and 8), and planned comparison revealed that, for words, there was a main effect for prosody (F(3,33) = 8.283, p = .000). Pairwise comparisons revealed that there was a significant difference in MMNs elicited between standard and pitch deviant (p=.049), and between standard and pitch+intensity deviant (p=.000), and intensity deviant and pitch+intensity deviant (p = .044).

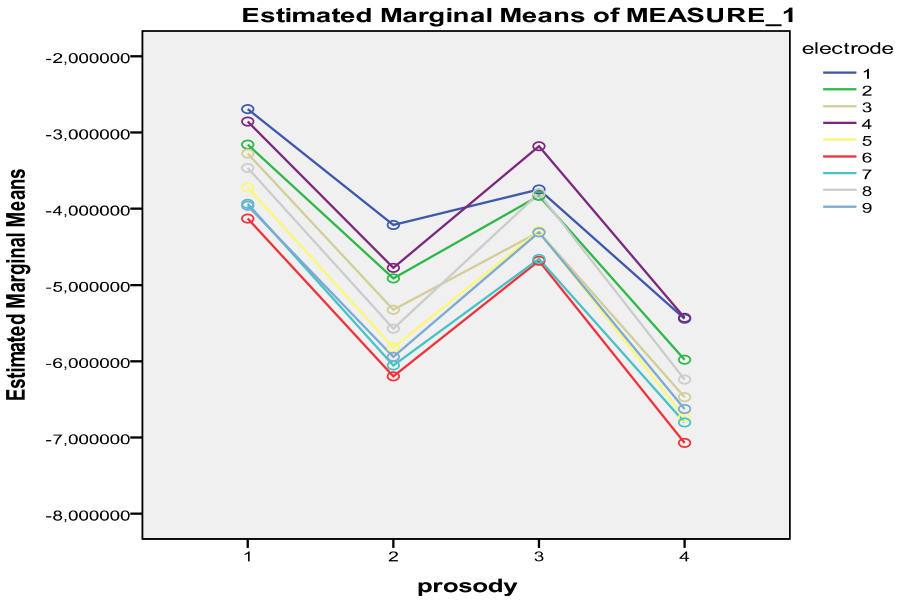


Figure 7 MMN differences for levels of the prosody in word block

For pseudowords, planned comparison revealed a main effect for the prosody (Word: F(3,33) = 9.682, p = .000). Pairwise comparisons revealed that there was a significant difference in MMNs elicited between the standard and pitch deviant (p=.000), and between the standard and pitch+intensity deviant (p=.001).

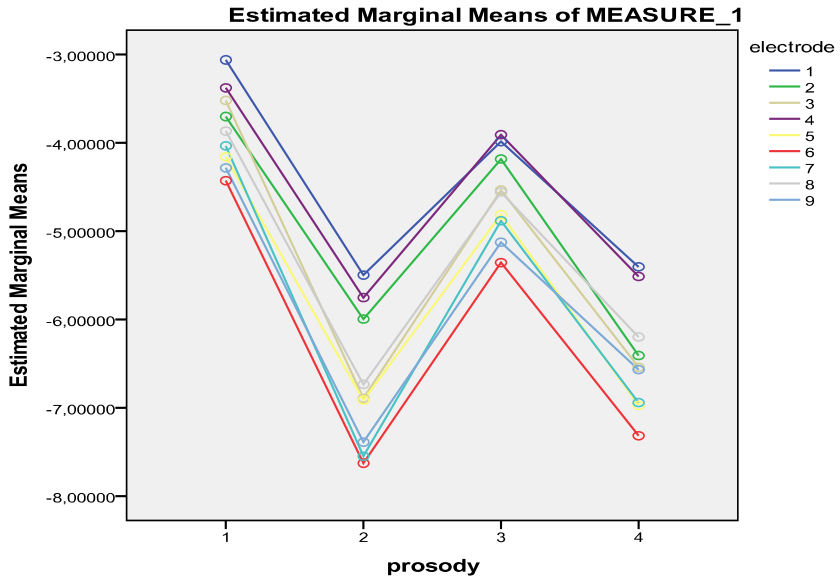


Figure 8 MMN differences for levels of the prosody in pseudoword block

6. Discussions

The objective of the analysis using ERP data was to characterize amplitude differences in the brain's electrophysiological response to different acoustic cues in lexical stress perception. Event-related potentials (ERPs) were recorded to series of frequent (standard) and three types of rare (deviant) words and pseudowords differing from the standard in frequency, intensity, or in both features. All deviants elicited the mismatch negativity (MMN), and the first MMN increment was around ~110, thereby supporting the first hypothesis which claimed that MMN elicitation should be realized as early as ~110 ms, indexing the deviation of a sound from the transient auditory memory trace of the standard.

Unlike the previous MMN studies in spoken word recognition, there was not a significant difference between MMNs elicited for words and pseudowords. Therefore the second hypothesis, concerning the role of prior knowledge of lexical stress in word recognition speed, was rejected. By failing to show a significant MMN amplitude difference between words and pseudowords in the first time window, the experiment reported in this paper provided no evidence that prior knowledge of prosodic structure can be used to speed word identification. Thus these results are at odds with a model of lexical access in which the lexicon can be partitioned by stress pattern in such a way that if stress pattern is given, the number of potential word candidates is reduced.

The results do not support the acoustic cues as an indicator of spoken word recognition in verb-noun pairs. A possible reason for this finding may be found in a disagreement about the role of suprasegmental cues in lexical access. Experimental evidence supports this, by suggesting that lexical access does not employ information accessible from English word prosody. In other words, listeners did not make a distinction between these two word forms when initially achieving access to the lexicon (McQueen & Cutler, 1997:581). Indeed, it is plausible that the listener does not make early use of lexical stress for lexical access. In order to perceive the stress pattern of a word, the listener's word recognition system should identify how many syllables there are in the word; in fact, therefore, it cannot initiate the process of lexical access until the end or nearly the end of the word (McQueen & Cutler, 1997:582). As stated by Lehiste (1970:1), suprasegmental features are detected by comparing two or more adjacent segments. For example, a vowel cannot be identified as stressed or unstressed without comparison to other vowels. Perhaps, therefore, prosodic cues do not take part in the pre-lexical access in English since the information they supplies cannot outweigh the drawback of delayed initiation of access.

The third hypothesis predicted that the stress shift from second to the first syllable should elicit higher MMN in words since in both cases, the result will be a meaningful word, i.e. hearing the noun rather than the verb. This hypothesis was tested by looking at the mean amplitude differences of MMNs elicited by word and pseudoword deviants in the second time window. The differences were not statistically significant. At the present time, we offer no conclusive explanation for why the hypothesis did not sustain statistical support. In accordance with the explanation offered for the rejection of the second hypothesis, the listeners should have been able to make use of lexical stress for lexical access because they had enough latency to compare the two syllables. Since MMN is a useful measure for comparison processes in the memory, and indexes automatic access to the items in the mental lexicon, it should have shown the differences between words and pseudowords if prosody participates in lexical access. A potential reason why the current study fell short in achieving the hypothesis is speculative.

The fourth hypothesis, concerning the relative importance of different acoustic cues, was confirmed. It was hypothesized that the most important cue should elicit the highest MMN, and in the light of the earlier research, fundamental frequency and the combination of fundamental frequency and intensity should elicit higher MMNs than intensity should. Statistical analysis demonstrated that while there were significant differences between standard and pitch deviant and pitch+intensity deviant, there was not a significant difference between standard and intensity deviant. As stated by Lehiste (1970:110), there is no one-to-one correspondence between stress and any single acoustic parameter. In the light of this statement, it can be asserted that the combination of two acoustic cues will ease the listeners perception and this was the case in the analysis of the present study. Furthermore, statistical results revealed another fact regarding the role of intensity. Pairwise comparisons showed that there was a significant difference between intensity deviant and pitch+intensity deviant in the second time window. Correspondingly, this indicated that while intensity, on its own, was not an important cue, it became prominent when combined with pitch.

7. Limitations of the current study

There are a number of limitations to the current study that should be taken into consideration when considering the results. The most notable reason for the limitations lies in the fact that the experimental design, data collection and analysis were to be completed in a limited time frame. The primary limitation of the current study is the small number of participants, limiting

the interpretation of the results and reducing the statistical power needed to detect significant interaction effects. Although attempts were made to obtain a larger sample size, it was difficult to recruit native speakers for participation in the EEG study. Another source of weakness in the current study is related to the limited number of trials. In particular, the small number of deviant trials placed great limitations upon this study, suggesting a lower level of evidence and statistical power that would have occurred with a larger number of trials. Another limitation is the fact that we cannot guarantee that the experimental material chosen was an optimal representative of the stress shift between noun-verb pairs and this may restrict the generalization of the findings. The same experiment which would be conducted with other materials may give different results. Finally, another caveat which needs to be noted regarding the present study is related to the manipulation of the experimental stimuli. Although we tried to create natural sounding synthetisized words, we cannot guarantee the naturalness of the real words, and accordingly, cannot generalize the findings for the stress judgments of participants in natural speech.

8. Further research

In light of the general findings presented in this study, it would be interesting to scrutinize whether acoustic cues have a similar influence on spoken-word recognition in languages other than English. In the future, a replication of this study with native speakers of Turkish could be realized. This will allow for comparison of word stress correlates perception across two different languages, belonging to different rhythmical classes. Turkish and English differ with respect to linguistic rhythm (English as a stress-timed language and Turkish as a syllable-timed language), and storage of stress patterns in the mental lexicon can cause cross-language differences in perceptual sensitivity to rhythmic parameters. Ingram noted that "Diffrences in prosodic systems produce significant prosodic interference effect (interference of L1 prosody on L2) for second language learners." (p.26). As it has been proposed by Suranyi, et al. (2009), sensitivity to the parameters underlying speech rhythm is significant in establishing well-specified phonological representations in the mental lexicon. However, different acoustic parameters can contribute differentially to rhythm and stress in different languages. So a follow-up study can be done to examine the perception of English stress by Turkish speakers to determine whether the Turkish prosody (L1) affects the prosody of English (L2).

9. Conclusions

The present investigation attempted to obtain additional information regarding the effect of prosodic cues on automatic word processing in the brain by comparing the mismatch negativity (MMN) component of the event-related potentials (ERP) elicited by isolated words and pseudowords. The findings of the present research confirmed the brain's automatic reaction to any change in the auditory sensory input. It has been sustained that the automatic process that compares incoming stimuli to a sensory memory trace of preceding stimuli can be demonstrated through MMN increments.

Moreover, MMN was established as a probe into auditory sensory memory and automatic change detection, by clarifying its dependence on particular physical stimulus characteristics. MMNs to changes in two basic stimulus features: fundamental frequency and intensity were addressed. The importance of fundamental frequency was confirmed by the data presented here. It has been shown that in English, changes in fundamental frequency can swing listeners perception of strong stress from the second to the first syllable. The data showed no case in which change of intensity ratio, on its own, caused a complete shift of the stress judgment from the second to the first syllable.

The present study provided no evidence for memory trace hypothesis which states that prior knowledge of prosodic structure can be used to speed word identification. The results obtained do not support the acoustic cues as an indicator of spoken word recognition in verb-noun pairs. A potential reason why the present investigation fell short in fulfilling the hypothesis is speculative. Further research should be undertaken in order to evaluate these findings and understand the nature of the spoken word processing in the brain.

References

Atienza, M., Cantero, J.L., Gomez, C.M.(1997). The mismatch negativity component reveals the sensory memory during REM sleep in humans. *Neurosci. Lett*., 237, 21–24.

Bond, Z. S., & Small, L. H. (1983). Voicing, vowel, and stress mispronunciations in continuous speech. *Perception & Psychophysics, 34*, 470–474.

Ceponeine, R., Kushnerenko, E., Fellman V., Renlund, M., Suominen, K., & Näätänen, R. (2002). Event-related potential features indexing central auditory discrimination by newborns. *Cognitive Brain Research,* 13, 101-113.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of Lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, 45 (3), 207-228.

Cutler A (2005). Lexical stress. In DB Pisoni & RE Remez (Eds.), *The handbook of speech perception*, pp. 264-89. Blackwell, Oxford.

Cutler, A., & Otake, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of Memory and Language,* 46, 296-322.

Cutler, A., & Norris, D. (1986). The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology, 14, 113-121.

Cutler, A., & Clifton, C. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183–196). Hillsdale, NJ: Lawrence Earlbaum Associates. Retrieved from: http://repository.ubn.ru.nl

Fear, B. D., Cutler, E. A., & Butterfield, S. (1995). The strong/weks syllable distinction in English. *The Journal of the Acoustical Society of America*, 97(3), 1893-1904.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The*

Journal of the Acoustical Society of America, 27 (4), 765-768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1(2), 126-152.

Greenberg, S. (2006). A multi-tier framework for understanding spoken language. In S. Greenberg & W. Ainsworth (Eds.), *Listening to speech – An auditory perspective* (pp. 1-32). Hillsdale, NJ: Lawrence Erlbaum Associates.

Honbolygo, F., Csepe, V., & Rago, A. (2004). Suprasegmental speech cues are automatically processed by the human brain: a mismatch negativity study. *Neuroscience Letters,* 363, 84–88

Ingram, J. C. (2007). *Neurolinguistics.* Cambridge: Cambridge University Press.

Jacobsen, T., Schröger, E. & Sussman E. (2004). Pre-attentive categorization of vowel formant structure in complex tones. *Cognitive Brain Research*, 20, 473-479.

Ladefoged, P. & Johnson, K. (2011). *A course in phonetics* (6th ed.). Boston, MA: Wadsforth. Retrieved from: http://library.nu/

Lehiste, I., & Fox, R. A. (1992). Perception of prominence by Estonian and English listeners. *Language and Speech*, 35(4), 419-434.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge Mass.: M.I.T Press.

McQueen, J. M., Cutler, A., & Norris, D. (2003). Flow of information in the spoken word recognition system. *Speech Communication*, 41, 257-270.

McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible word-constraints in Japanese speech segmentation. *Journal of Memory and Language* 45, 103-132.

McQueen, J. M. & Cutler, A. (1997). Cognitive processes in speech perception. In W. J. Hardcastle and J. Laver (Eds.), *A Handbook of Phonetic Science* (pp. 566-585). Oxford:

Blackwell.

Morton, J., & Jassem, W. (1965). Acoustic correlates of stress. *Language and speech*, 8, 159-181.

Moss, H. E., & Gaskell, M. G. (1999). Lexical semantic processing during speech comprehension. In S. Garrod & M. Pickering (Eds.), *Language Processing* (pp. 59-99). Hove, UK: Psychology Press.

Nelson, C. (1998). *Attention and Memory: An integrated framework*. New York: Oxford University Press.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38, 1-21.

Näätänen, R. (1999). Phoneme representations of the human brain as reflected by event-related potentials. *Electroencephalography and Clinical Neurophysiology: Supplement*, 49, 170–173.

Näätänen R & Winkler I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull*, 125, 826–59.

Näätänän, R. (1992). *Attention and brain function*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Näätänen, R., Gaillard, A. W. K. & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica,* 42, 313-329.

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology,* 79, 49–71.

Pulvermüller, F., Shtyrov, Y., Kujala, T. & Näätänen, R. (2004). Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology* 41, 106-112.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R.J., Näätänen, R., (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage* 14, 607–616

Roll, M. (2009). *The neurophysiology of grammatical constraints*. Doctoral dissertation. Lund University.

Sams, M., Paavilainen, P., Alho, K., and Näätänen, R., Auditory Frequency Discrimination and Event-related Potentials, *Electroencephalogr. Clin. Neurophysiol.*, 62, 437- 448.

Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. Behaviour Research Methods, *Instruments & Computers*, 30(1), 131-145.

Shtyrov, Y., Kujala, T., & Pulvermüller, F. (2009). Interactions between Language and Attention Systems: Early Automatic Lexical Processing? *Journal of Cognitive Neuroscience*, 22(7), 1465-1478.

Shtyrov, Y., & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport, 13*(4), 521–525.

Slowiaczek, L., M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, 33(1), 47-68.

Soto-Faraco, S., Sebastian-Galles, N., & Cutler, A. (2001). Segemental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*. 45, 412-432.

Suranyi, Z., Csepe, V., Richardson, U., Thomson, J. M., Honbolygo, F. & Goswami, U. (2009). Sensitivity to rhythmic parameters in dyslexic children: a comparison of Hungarian and English. *Read Writ* 22, 41-56.

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables.

Acoust. Soc. Am., 89, 1768-1776.

Ylinen, S., Strelnikov, K., Huotilainen, M., & Näätänen, R. (2009). Effects of prosodic familiarity on the automatic processing of words in the human brain. *International Journal of Psychophysiology 73,* 362–368.

Weber, C., Hahne, A., Friedrich, M. & Friederici, D. (2004). Discrimination of word stress in early infant perception: electrophysiological evidence. *Cognitive Brain Research* 18, 149-161.

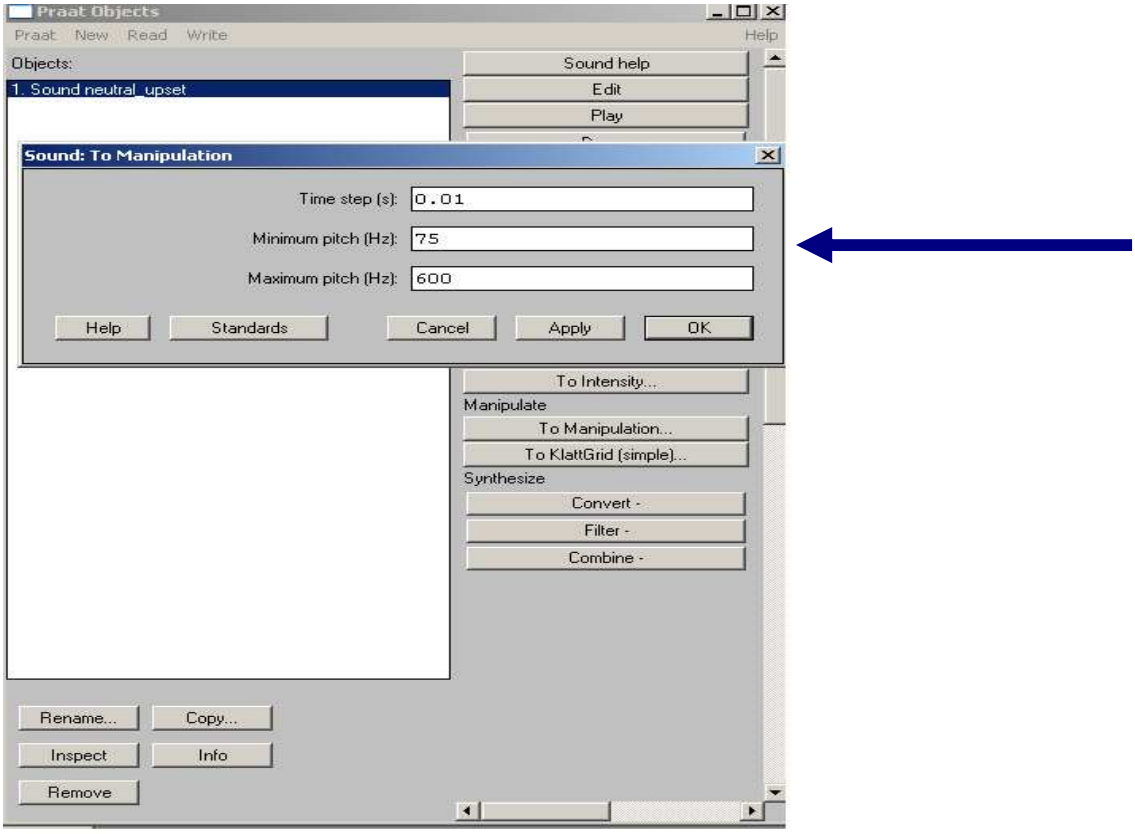
Zsuzsanna, S., Csepe, V., Richardson, U., Thomson J. M., Honbolygo, F., & Goswami, U. (2008). Sensitivity to rhythmic parameters in dyslexic children: a comparison of Hungarian and English. *Springer Science+Business Media B.V*.

APPENDICES

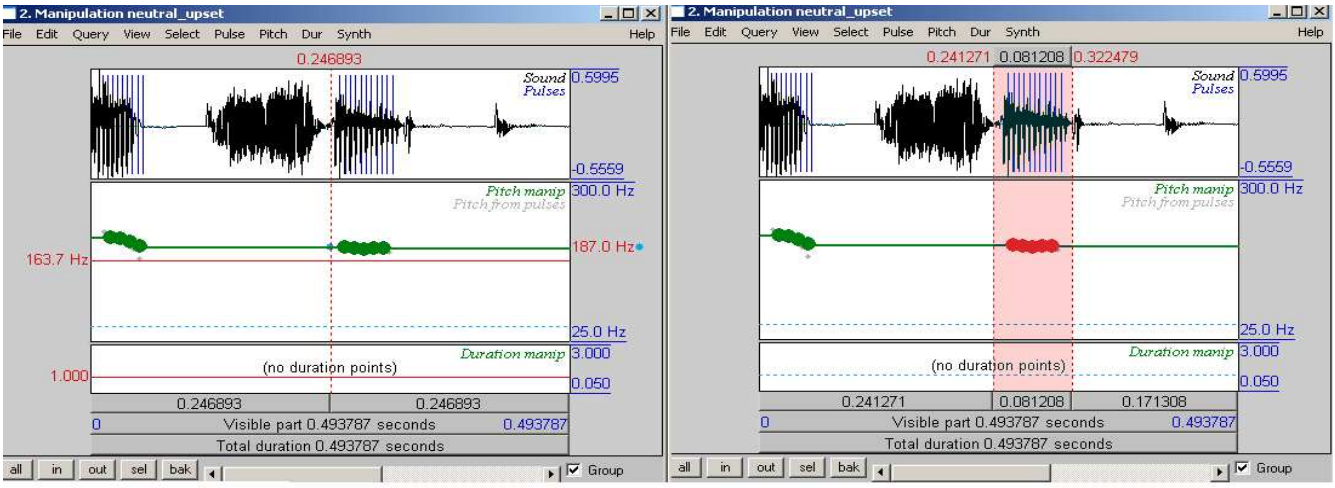
Appendix A – Manipulation of pitch, duration, intensity using PSOLA implemented in Praat

Pitch resynthesis

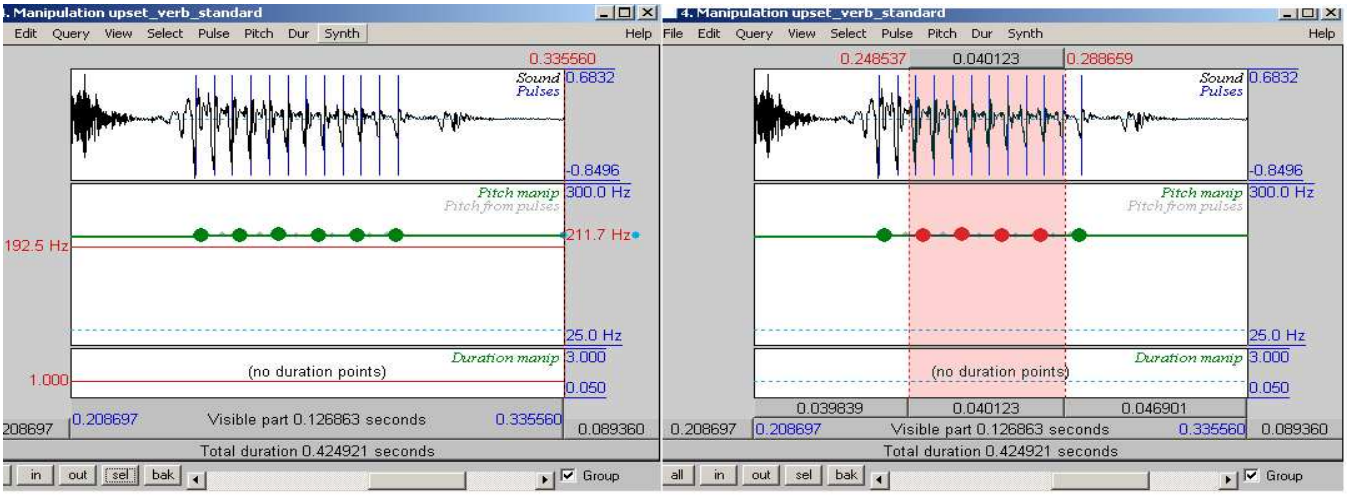
1. Select the source sound and create a manipulation object from it: (button) **To Manipulation**



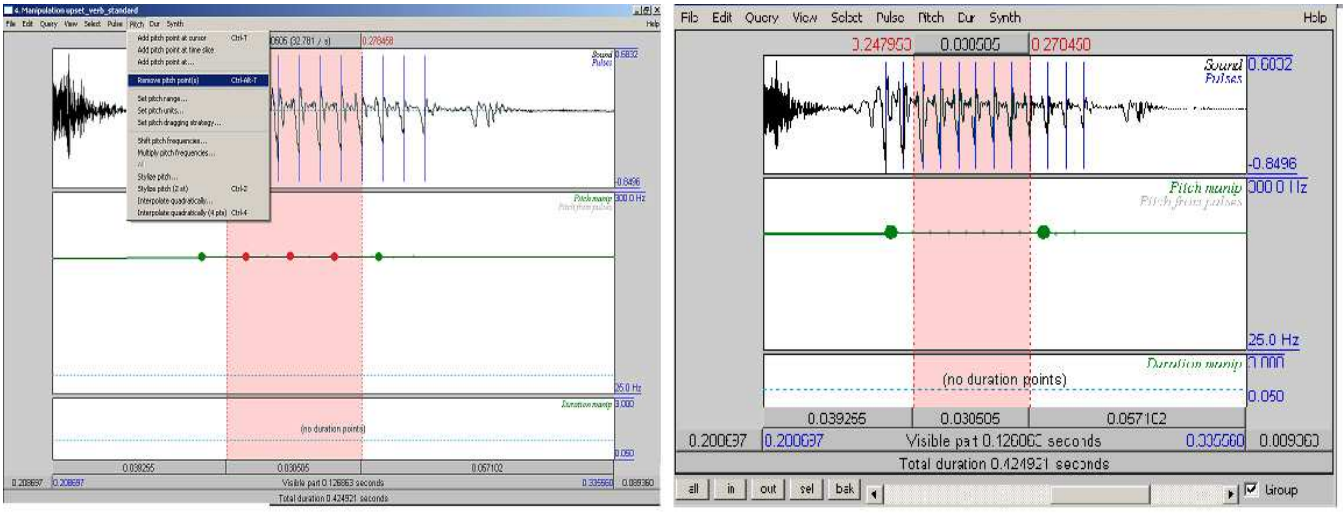
2. With the manipulation object selected, open the manipulation window: (button) **Edit**
   Select a particular segment that you want to change the pitch contor and zoom in to it: (button) **sel**.



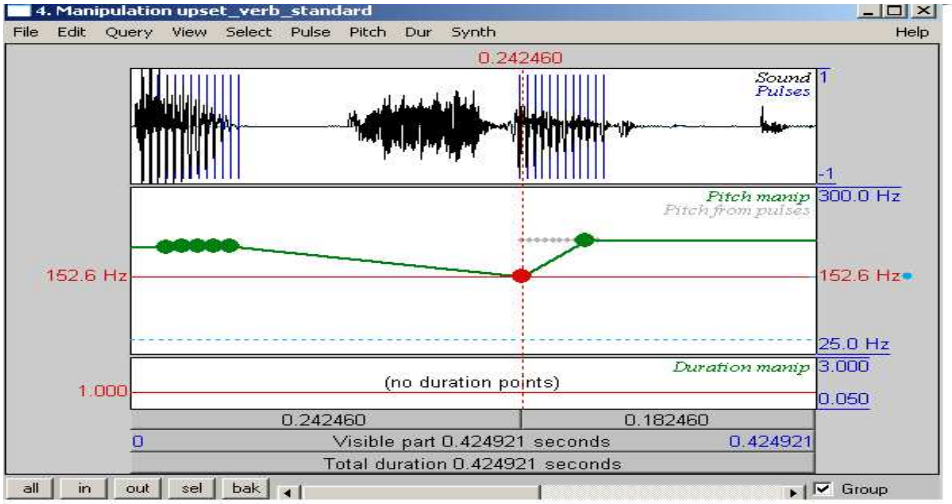
To stylize the pitch contor of a particular segment, you would first need to determine which points are perceptually important.



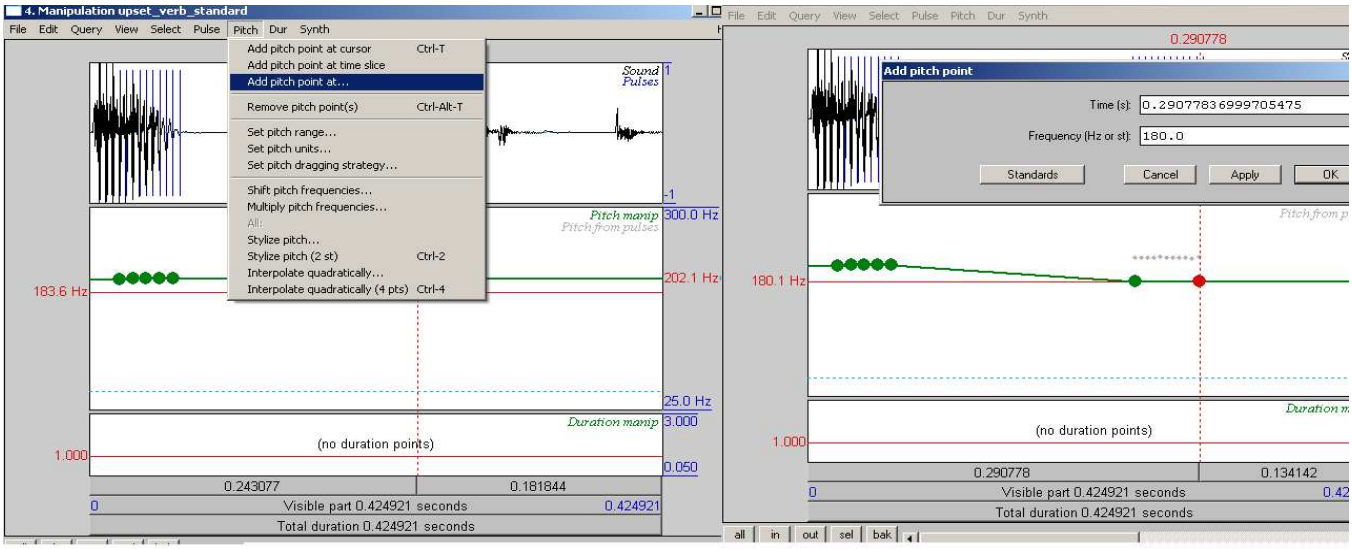
Once the anchoring pitch points are determined, you can safely remove the other pitch points: (menu) **Pitch** ==> **Remove pitch point(s)**



3. Holding the pitch point with your left-mouse-button, **drag it around** to set a new f0 value to that point.



4. However, if you want to be exact in your stylization, you can use the command '**Add pitch point at...**' and define the exact pitch value in the time point that you want.



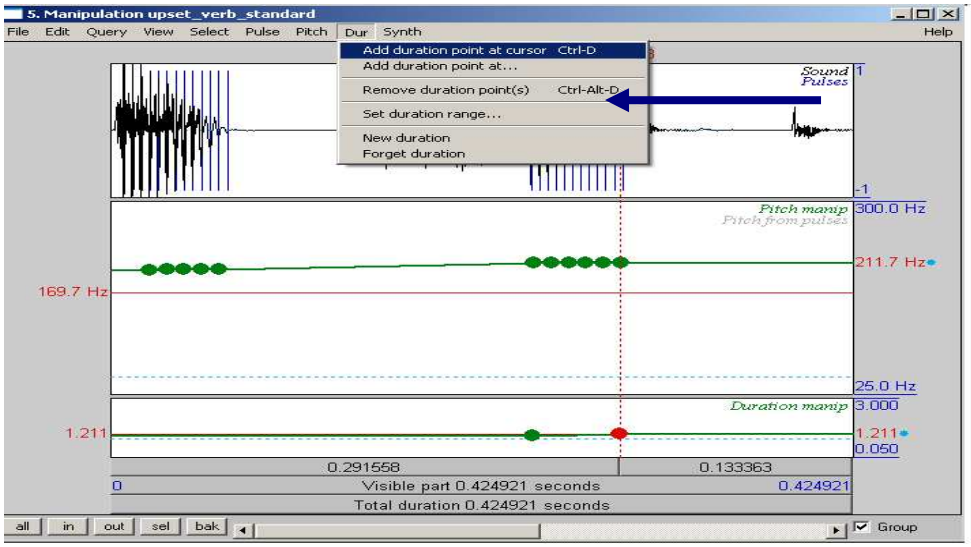
5. Play the resynthesized object to listen to the new pitch contour. When you're satisfied with what you hear, you can create as a sound object the resynthesized version of your stylized pitch contour: You can do this in the manipulation window: (menu) **File** ==> **Publish resynthesis**.

Duration Resynthesis

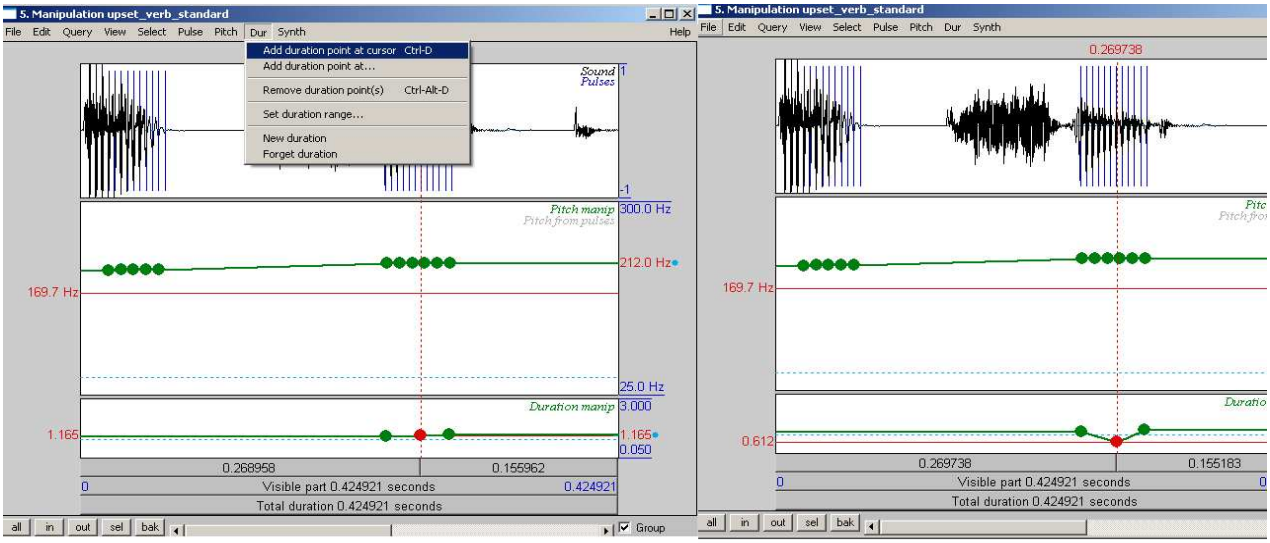
Steps 1 through 3 are the same as in the pitch contour manipulation.

4. If you want to modify a particular section of an object, you need at least three or more duration points. With three duratoin points, you can "gradually" increase/decrease the duration of the vowel segment. First, insert two duration points at the beginning and end of the vowel segment by selecting the corresponding "pitch points". You do this by first selecting the beginning pitch point of the vowel segment by clicking on the pitch

point, followed by (menu) **Dur** ==> **Add duration point at cursor**. When a duration point is inserted, it is set to the factor of 1 by default (as indicated by the red 1.000 on the y-axis of the window)



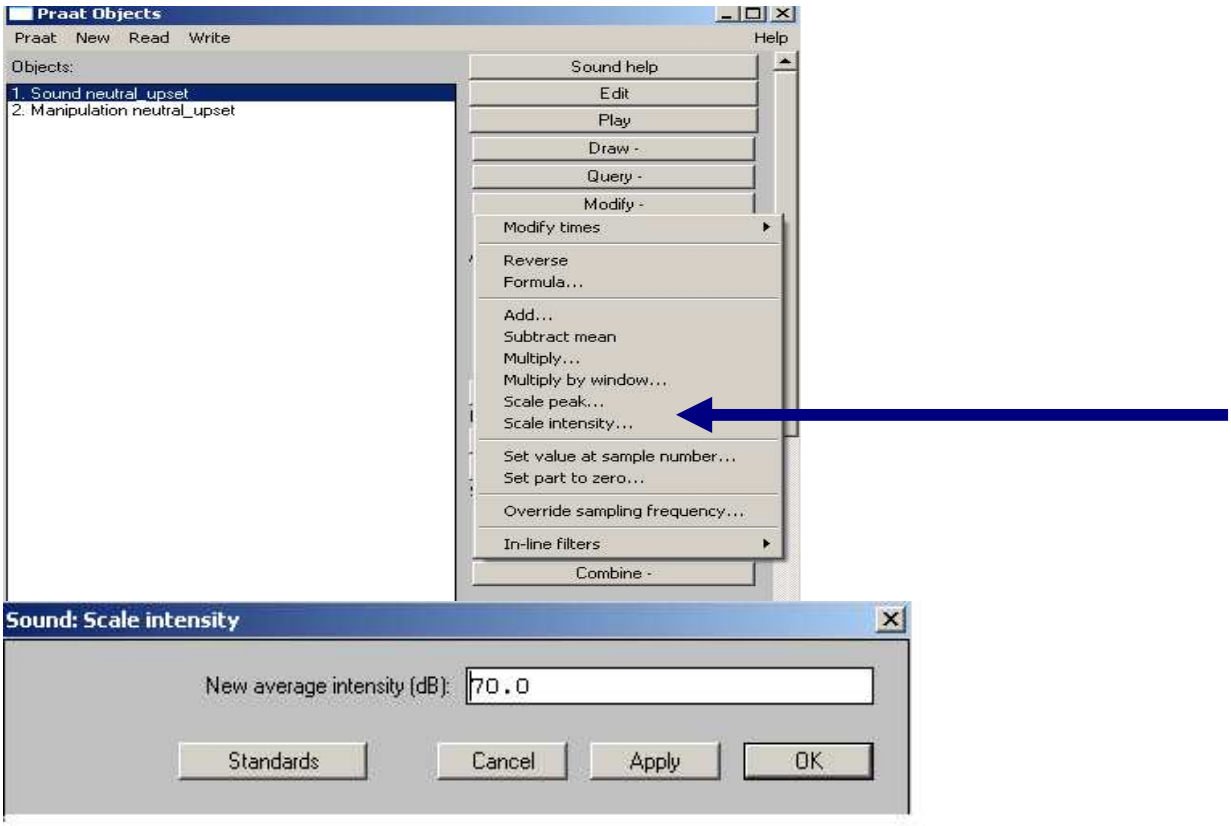
6. Insert a third duration point somewhere in between the two duration points. Click-hold the mid duration point and drag it around to manipulate the duration of the vowel segment



7. If you want, you can create a resynthesized object: (menu) **File** ==> **Publish resynthesis**. Then you can locate this object in the object window (below) and it is named "fromManipulationEditor".

Intensity resynthesis

To increase or decrease overall intensity of a particular segment, use the command '**Scale intensity...**' and then write the new value you would like to implement.



Appendix B – Recording script

upset /ʌp'sɛt/ (verb): make (someone) unhappy, disappointed, or worried

upset /'ʌp.sɛt/ (noun): an unexpected result or situation:

---

impact /ɪm'pækt/ (verb): have a strong effect on someone or something

impact /'ɪm.pækt/ (noun): a marked effect or influence:

---

increase /ɪn'kris/ (verb): become or make greater in size, amount, or degree

increase /'ɪn.kris/ (noun): a rise in the size, amount, or degree of something

---

transport /træns'pɔrt/ (verb): take or carry (people or goods) from one place to another

transport /'træns.pɔrt/ (noun): the action of transporting something or being transported

---

imprint /ɪm'prɪnt/ (verb): to mark a surface by pressing something hard into it

imprint /'ɪm.prɪnt/ (noun): any impression or impressed effect

---

replay /ri'pleɪ/ (verb): to play back (a recording on tape, video, or film)

replay /'ri.pleɪ/ (noun): the playing again of part of a recording

1st recording

UPset (noun) UPset (noun)
upSET (verb) upSET (verb)

IMpact (noun) IMpact (noun)
imPACT (verb) imPACT (verb)

INcrease (noun) INcrease (noun)
inCREASE (verb) inCREASE (verb)

TRANSport (noun) TRANSport (noun)
transPORT (verb) transPORT (verb)

IMprint (noun) IMprint (noun)
imPRINT (verb) imPRINT (verb)

REplay (noun) REplay (noun)
rePLAY (verb) rePLAY (verb)

2nd recording

Verb: The bad news upset me. **Upset**
Noun: She didn't realize the upset she caused me.**Upset**

Verb: This decision may impact your career. **Impact**
Noun: This study explores the impact of stress. **Impact**

Verb: They want to increase taxes on Americans. **Increase**

Noun: There has been an increase in the number of studies about prosody. **Increase**

Verb: The ship will transport the cargo quickly. **Transport**

Noun: The cost of transport will outweigh our savings. **Transport**

Verb: We are going to imprint the neural patterns of the brain onto a computer. **Imprint**

Noun: Each molecule carries the imprint of its ancestors. **Imprint**

Verb: He could replay the tape whenever he wished. **Replay**

Noun: I watched the replay a couple of times. **Replay**

3rd recording

I said **UPset**. I said **UPset** this time.

I said **upSET**. I said **upSET** this time

I said **IMpact**. I said **IMpact** this time.

I said **imPACT**. I said **imPACT** this time.

I said **INcrease**. I said **INcrease** this time.

I said **inCREASE**. I said **inCREASE** this time.

I said **TRANSport**. I said **TRANSport** this time.

I said **transPORT**. I said **transPORT** this time.

I said **IMprint**. I said **IMprint** this time.

I said **imPRINT**. I said **imPRINT** this time

I said **REplay**. I said **REplay** time.

I said **rePLAY** . I said **rePLAY** time