1-1-2004

The impact of dynamic changes in talker amplitude on recognition memory for words

Kimberly M Cramer
University of Nevada, Las Vegas

THE IMPACT OF DYNAMIC CHANGES IN TALKER AMPLITUDE ON RECOGNITION MEMORY FOR WORDS

by

Kimberly M. Cramer

Bachelor of Science
University of Nevada, Las Vegas
1997

Master of Science
University of Nevada, Las Vegas
2001

A thesis submitted in partial fulfillment
of the requirements for the

Master of Arts Degree in Psychology
Department of Psychology
College of Liberal Arts

Graduate College
University of Nevada, Las Vegas
May 2005

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

Thesis Approval
The Graduate College
University of Nevada, Las Vegas

April 14 , 20 05

The Thesis prepared by

Kimberly M. Cramer

Entitled

The Impact of Dynamic Changes in Talker Amplitude on Recognition Memory for Words

is approved in partial fulfillment of the requirements for the degree of

Master of Arts in Psychology

Examination Committee Chair

Dean of the Graduate College

Examination Committee Member

Examination Committee Member

Graduate College Faculty Representative

1017-53

ABSTRACT

The Impact of Dynamic Changes in Talker Amplitude on Recognition Memory for Words

by

Kimberly M. Cramer

Dr. Michael Hall Examination Committee Chair
Assistant Professor of Psychology
University of Nevada, Las Vegas

This study investigated whether dynamic changes in the amplitude of speech were represented along with word information. An emotional manipulation was used to examine if listeners were sensitive to dynamic changes in amplitude. In Experiment 1, six talkers produced 200 phonetically balanced (PB) words with different intended emotions (e.g., joy versus sadness). Intensity measurements across time were recorded for each target word. Statistically distinct amplitude contours were obtained as a function of intended emotion. In Experiment 2, listeners judged whether each word in a list of spoken words was "new" (i.e., word was new to the list) or "old" (i.e., word was presented earlier in the word list). Listeners were more accurate at recognizing a word as old if it was repeated by the same talker; however, there was no recognition advantage for words repeated in the same amplitude contour. In Experiment 3, listeners were asked to discriminate joy from sad amplitude contours imposed on consonant vowel (CV) syllables within two versions of a two-alternative forced choice task (unequated versus equated). Listeners' sensitivity to contour differences was high in the unequated version

of the task and poor in the equated version indicating that listeners primarily rely on overall loudness differences to discern between joy and sad contours. Implications for the potential role of amplitude in the perception of emotion, and for the representation of talker information, are discussed.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Everyday we are inundated by the speech of multiple-talkers, each producing utterances with a variety of idiosyncratic differences. Despite this enormous range of talker variability, listeners readily perceive utterances across talkers, without necessarily having conscious awareness of the source characteristics. The ability to extract stable linguistic percepts (i.e., words and phonemes) from an acoustic speech signal that varies substantially is referred to by speech researchers as the problem of perceptual constancy (Verbrugge, Strange, Shankweiler, & Edman, 1976). For example, the utterance, "The girl kicked the ball." is recognized similarly whether it is spoken by a friend or by a complete stranger. This general observation has traditionally led speech researchers to assume that there might be a perceptual process or mechanism that automatically "normalizes" acoustic differences among talkers in order to preserve perceptual constancy and abstract the canonical units comprising the linguistic message (Mullenix, Pisoni, & Martin, 1989). For instance, the phonemes that comprise the words, "The girl kicked the ball." constitute the fundamental canonical units of the message.

CHAPTER 2

REVIEW OF RELATED LITERATURE

A growing body of evidence alternatively suggests that the processes of speech perception and spoken word recognition may operate in the context of highly detailed, instance specific representations of the acoustic signal (see Goldinger, Pisoni, & Logan, 1991). Recent investigations involving the identification and recognition of spoken words demonstrate that experience with a talker facilitates performance on these tasks (Bradlow, Nygaard, & Pisoni, 1999; Church & Schacter, 1994; Goldinger et al., 1991; Mullenix & Pisoni, 1990; Mullenix et al., 1989; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1995; Nygaard, Sommers, & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993; Schacter & Church, 1992; Sommers et al., 1994). Moreover, these studies indicate that linguistic content and a talker's voice are processed together in an integral fashion (Mullenix et al., 1989; Mullenix & Pisoni, 1990). This research also has shown that talker variability affects the accuracy of recall of spoken words not only by increasing the processing demands for early perceptual encoding of the words, but also by affecting the efficiency of the rehearsal process itself (Goldinger et al., 1991).

Closer examinations of talker effects reveal that listeners are sensitive to at least two aspects of a talker's voice that consequently must be encoded and represented with word information[1]. These vocal characteristics include the rate at which words are produced by a given talker (e.g., Bradlow et al., 1999; Martin, Mullenix, Pisoni,

&Summers, 1989; Mullenix et al., 1989; Mullenix & Pisoni, 1990; Palmeri et al., 1993; Nygaard et al., 1995; Sommers et al., 1994) and gender (Palmeri et al., 1993), in particular, fundamental frequency, or F0 (Church & Schacter, 1994; Schacter & Church, 1992). In contrast, researchers have failed to find an influence of amplitude on listeners' performance (see Bradlow et al., 1999; Church & Schacter, 1994; Nygaard et al., 1995; Sommers et al., 1994). The lack of findings for an amplitude effect has led some researchers (see Bradlow et al., 1999) to conclude that listeners do not represent amplitude with word information because it does not contain additional information that would be helpful to the listener. For example, Bradlow et al. (1999) have suggested that amplitude does not convey linguistically relevant information that would facilitate extraction of the meaning from the utterance.

One potential reason for the lack of amplitude effects is due to the fact that researchers have only studied a static aspect of amplitude. That is, researchers have manipulated overall amplitude by scaling the signal presentation levels up or down across a decibel (dB) range. Thus, amplitude variability was a constant one-dimensional adjustment. Perhaps amplitude is coded as a dynamic attribute, such as an amplitude contour, which represents fluctuations in intensity across the production. For example, it may be that a talker's emotional state during the production of words is encoded as momentary changes in amplitude. In an effort to learn more about the underlying perceptual and memory processes involved in the recognition of spoken words, the present investigation seeks to examine whether manipulating the amplitude contour of a signal produced in a specific emotion significantly impacts spoken word recognition.

Talker Characteristics

Differences in voice characteristics among individual talkers are due to a wide variety of factors. For example, structural factors related to the physical shape and length of the oral and nasal cavities in the vocal tract constrain the ultimate acoustic composition of the speech signal (Fant, 1960). One way that this may be illustrated is by considering the differences in the vocal tract size among men, women, and children, and how these differences affect the formant frequencies of vowels. A formant is a concentration of acoustic energy in a band of frequencies during speech production. There are several formants, each with a different center frequency, and each corresponding to a resonance mode in the vocal tract. These structural differences result in variations in voice characteristics between talkers. One consequence of differences in vocal tract size and shape is that the acoustic properties of vowels may vary substantially (Peterson & Barney, 1952). For example, Peterson and Barney (1952) found that the first and second formant frequencies for vowels produced by women are higher compared to those produced by men, and are even higher for those produced by children. Differences in glottal source function also exist between talkers, resulting in other voice quality differences that distinguish speakers (see Carrell, 1984).

In addition to anatomical or structural factors, a number of more dynamic factors affect the speech signal, such as the control and positioning of the articulators and the manner in which the vocal gestures are executed (Ladefoged, 1980). For example, lip rounding, such as that used in production of the word "book", acts to lengthen the vocal tract and reduces the first and second formant frequencies of a word. Given the

substantial acoustic differences between talkers, the problem of managing these sources of variability in perception becomes an important research issue.

Theories of Spoken Word Recognition and Representation

Talker variability has typically been considered a "perceptual problem" to be solved by listeners, just as it must be solved in the design of adequate speech-recognition systems (Goldinger et al., 1991). The traditional abstractionist approach to explaining how listeners contend with talker variability in speech is through a compensatory process, in which speech sounds are normalized with reference to a talker's voice. A strict interpretation of this view is that surface information about a talker is "stripped" away during the perception of speech because it is linguistically irrelevant. According to this view, the purpose of normalization is to produce a standardized, abstract, linguistic-phonetic representation, which can then be matched to stored canonical forms (Nygaard, Sommers, & Pisoni, 1994; Rand, 1971; Summerfield, 1975).

Typically, normalization researchers have evoked two approaches to study normalization processes in speech perception. One approach has been to search for acoustic-articulatory invariants believed to permit access to phoneme and word-sized units (e.g., Joos, 1948). The other approach has been to focus on normalization algorithms and processes that filter out stimulus variability to reach the core, abstract units argued to underlie linguistic processing (e.g., Sussman, 1986).

The abstractionist approach, with its emphasis on context-free processing units, fails to provide a satisfactory explanation for the relationship between the processing of linguistic information and the analysis of a talker's voice. More recent studies that have

explicitly examined the effect of talker, along with more specific voice parameters, on memory for spoken words, clearly demonstrate that experience with a talker improves recognition performance for spoken words. For example, Nygaard et al. (1994) and Nygaard and Pisoni (1998) demonstrated that with training, listeners familiarized with a set of talkers showed continuous improvement in their ability to recognize talkers from isolated words. These researchers suggested that because the speech processing system can be modified with training and experience with a talker's voice, specific voice characteristics are encoded into memory. Furthermore, they contended that these stored characteristics are used to facilitate the perceptual analysis of words spoken by a particular talker.

A separate body of research addressing the perception and identification of talker identity also calls into question the fundamental assumptions of the abstractionist approach (Bradlow et al., 1999; Nygaard et al., 1995). This research suggests that inherent talker variability in the speech signal is valuable information, utilized by the listener to facilitate the perceptual identification of spoken words. This alternative perspective assumes that variability is incorporated into lexical representations along with linguistic content (Bradlow et al., 1999; Goldinger, 1996; Pisoni, 1997). Linguistic representations are hypothesized to be extremely detailed, preserving the constantly changing surface form of each spoken word (Goldinger, 1996). In particular, this approach draws from exemplar-based theories of memory (Eich, 1982; Hintzman, 1986) and categorization (e.g., Nosofsky & Zaki, 1998). Episodic theories of the lexicon have been proposed that assume that collections of detailed exemplars represent individual words (Goldinger, 1996; Pisoni, 1997; Tenpenny, 1995).

Exemplar-based models assume that new information is categorized, or conversely is rejected, as an example of a particular category, based on feature similarity to the exemplars already represented within the category (Medin & Schaffer, 1978; Nosofsky, 1986). A new experience activates all stored exemplars, which may each be used in considering how to categorize the new experience. Hence, a new item is compared against all the stored exemplars that share similar features. The degree of feature overlap between the new item and stored items determines the perceptual distance between them. Exemplar theory predicts that a new instance will be identified as a member of the category that reflects the smallest average distance from stored exemplars. Likewise, classification accuracy and confidence are purported to increase for exemplars that are presented with high frequency. Conversely, for the classification of infrequent items that are dissimilar to the high-frequency exemplars, accuracy and confidence are posited to decrease.

Thus, according to exemplar-based approaches, every time a certain word is spoken, aspects of its acoustic attributes are accessed. These auditory patterns are labeled as the meaning of the word becomes known and also as a result of experience. The end product is a network of sound-based structures, each one linking a set of acoustic parameters to meaning (i.e., a semantically and grammatically defined category). In a strong form, episodic theories of speech perception predict that all aspects of surface form are included in lexical representation and that they affect both the identification and the recognition of spoken words (Bradlow et al., 1999).

Like exemplar-based models, prototype theory suggests that information is categorized and represented. However, unlike the exemplar account, the prototype theory

posits that categorization occurs by comparing a new item with a single abstract summary representation or "prototype" (Rosch, 1978). The prototype is an "ideal" image of the perfect member of a category. Representations are composed of ideal features or dimensional values for words in the category.

Prototypes can serve as the cognitive reference points against which actual items are judged during categorization. Therefore, words that are perceptually similar to the prototype are judged to be better instances of the category than items that are less similar to the prototype (Rosch, 1975). This suggests that members of a category form a gradient of typicality in which some items are more prototypical than others. For example, prototype theories are attractive to researchers working on theoretical issues in speech perception because prototypes provide one way of dealing with the lack of acoustic-phonetic invariance in the acoustic signal (Lively & Pisoni, 1997). Moreover, prototype theories attempt to solve the problem of perceptual invariance by assuming that listeners compare the incoming signal to an idealized form or "average" representation. Like exemplar theory, if the item is sufficiently similar to the prototype (i.e., more similar than to prototypes from other categories), then it is accepted as a member of a phonetic category. Thus, perceptual invariance is not found in the acoustic signal, but rather is achieved during the process of categorization by comparison with idealized forms or stored representations (Lotto, Kluender, & Holt, 1998).

Researchers have recognized that the prototype theory makes similar predictions to the exemplar-based approach. After all, the prototype for a given category may be determined from exemplars (e.g., average representation). Consequently, recent efforts by researchers have attempted to disentangle these similar but distinct views (e.g., Minda &

Smith, 2001; Nosofsky & Zaki, 2002; Palmeri & Nosofsky, 2001; Smith & Minda, 2001; Zaki & Nosofsky, 2001). For example, studies of spoken word recognition suggest that particular vocal parameters, such as rate and F0 are represented as individual features that are compared with future word productions. Although this evidence is consistent with an exemplar-based approach, it does not preclude the prototype theory because other parameters, such as amplitude have not been found to be represented. In each examination, the manipulation of amplitude was above levels required for reliable discrimination.

The failure to find amplitude effects argues against a true exemplar theory, which would predict encoding and retention of all aspects of surface form (Bradlow et al., 1999). Further examinations of amplitude are warranted before it can be regarded as a feature that is not represented. Prior examinations have only manipulated overall amplitude. It is possible that a static adjustment of overall amplitude, in which the level of dB is simply either increased or decreased, does not signal phonetic contrasts or add meaning. Rather than representing a static aspect of amplitude a word or utterance, it is possible that dynamic aspects of amplitude are represented. Like rate and F0, perhaps, another dimension of amplitude is represented that reflects differences in the complexity of the acoustic correlates.

Talker Variability Effects on Word Identification and Memory Performance

Several investigations have clearly demonstrated that talker variability can impact listeners' perceptual processing and recognition memory for spoken words. In experiments that have used multiple-versus-single-talker spoken word lists, decrements in

word identification and recognition memory performance have been observed (Mullenix et al., 1989). For example, Mullenix and Pisoni (1990) found that listeners had longer response latencies and poorer identification performance for words produced in multiple-talker contexts relative to words produced in single-talker contexts. Furthermore, these and other findings (e.g., Goldinger et al., 1993; Sommers et al., 1994) suggest that variability due to talker characteristics is both time and resource demanding. Hence, as a talker's voice changes from trial to trial in these tasks, listeners must devote additional processing resources to recover the phonetic content of the utterance. Martin et al. (1989) found that serial recall of spoken word lists produced by multiple-talkers was poorer than recall of lists produced by a single-talker in the primacy portion of the serial recall curve (i.e., items one through three in a ten-item word list). This led Martin et al. (1989) to conclude that recall of word lists by multiple-talkers demanded greater processing resources for the encoding and subsequent rehearsal of words in working memory than did the recall of single-talker lists. Furthermore, these authors found multiple-talker lists produced poorer recall of visually presented digits shown prior to the serial recall task. This finding was interpreted as providing further evidence that the encoding and rehearsal of words from multiple-talker lists required more processing resources than did the single-talker lists, leaving fewer resources available for storage and retrieval of the visually presented digits.

Goldinger et al. (1991) elaborated on Martin et al.'s (1989) findings to show that talker information could serve as a retrieval cue if listeners' are given a sufficient amount of time to process each talker's voice. Goldinger et al. (1991) varied the lag between the presentations of items from single-talker and multiple-talker spoken word lists. These

researchers observed differential effects for presentation rates. For fast presentation rates (one word every 250 ms), serial recall of spoken words was better in initial list positions for single-talker lists than for multiple-talker lists. In contrast, at slower presentation rates (one word every 4000 ms), this difference in recall accuracy was reversed. Goldinger and his colleagues interpreted these findings to suggest that talker variability affects not only early perceptual encoding, but rehearsal processes as well.

Other evidence indicates that a talker's voice and phonetic information are not perceptually independent; rather, they are processed together in an integral fashion. Mullenix and Pisoni (1990) employed a Garner (1974) speeded classification task, in which listeners were instructed to classify spoken words according to either phonetic identity (/b/ v. /p/) or talker gender (male v. female). The nature of the task required listeners to selectively ignore variation along the irrelevant dimension (either phoneme or talker). Mullenix and Pisoni observed an asymmetric pattern of interference between dimensions, which was reflected in slower reaction times (RT) and decreased identification accuracy. In addition, they found that the greatest interference was caused by irrelevant variation in the voice dimension. Furthermore, the analysis of phonetic information contained in initial consonants was more dependent on the prior or concurrent analysis of voice information. Thus, listeners had difficulty selectively ignoring voice information and attending only to the initial phoneme. This suggests that the processes involved in phonetic coding and the processes involved in encoding characteristics of a talker's voice do not operate independently.

Moreover, detailed talker-specific information is represented and facilitates spoken word recognition in both implicit and explicit memory tasks. For example, in a

continuous recognition memory procedure, an explicit memory task, Palmeri et al. (1993) examined listeners' word recognition performance for items that were repeated by the same and different talkers. Their findings indicated that listeners' recognition performance was better for words produced by the same talker. Listeners also were able to explicitly recognize whether or not the talker was the same or different as in the first occurrence of the word. Similarly, Schacter and Church (1992) and Church and Schacter (1994) found that a change in talker from study to test impaired participants' performance on an auditory stem completion task, an implicit measure of memory. Hence, words were more likely to be produced as a stem completion if the same talker at study and test repeated the stem. Goldinger (1996) also showed that words were identified and recognized more often when they were repeated by the same talker than when they were repeated by a different talker in tasks of perceptual identification (an implicit memory task) and recognition memory (an explicit memory task).

Finally, training studies, in which listeners are exposed to a set of ten talkers, have revealed that familiarity with a talker's voice increases intelligibility of words. Specifically, listeners who are given words produced by familiar talkers at test show better identification performance relative to listeners who are given words produced by unfamiliar talkers (Nygaard & Pisoni, 1998; Nygaard et al., 1994). Furthermore, Nygaard and Pisoni (1998) have demonstrated that generalization and transfer from voice learning to linguistic processing is sensitive to the talker-specific information available during learning and test. Hence, Nygaard and Pisoni (1998) found that learning a talker's voice from sentence-length utterances did not enhance or generalize well to identification of novel isolated words. They did, however, observe perceptual learning of novel voices

from sentence-length utterances improved speech intelligibility for words in sentences. Taken together, findings from overall training studies suggest that listeners appear to attend to specific dimensions of talker identity that are most relevant at test. These data provide support for the idea that talker characteristics are represented and can function as an effective retrieval cue.

Fundamental Frequency Effects

Studies of spoken word recognition indicate that the vocal parameter of F0 is represented and influences spoken word recognition (Church & Schacter, 1994; Palmeri et al., 1993; Schacter & Church, 1992). F0 is the primary acoustic correlate of vocal pitch (Lieberman, 1961). Subsequently, F0 represents the most obvious difference between the male and female voice. The average speaking F0 for male voices varies between 100 to132 hertz (Hz), whereas the average F0 for females varies between 142 to 256 Hz. The differentiation is because the mass of a male's vocal folds is greater, and the length of his resonating vocal tract is much longer (Coleman, 1971). In so far as different ranges of F0 should contribute to perception of a roughly corresponding distinct range of pitches, F0 should be a reliable indicator of a speaker's gender. Listeners may then use these higher or lower F0 values as a cue for gender identification.

Like the aforementioned general effects of a talker, there is converging evidence from both implicit and explicit memory tasks that suggests that F0 is a voice characteristic that is encoded and stored. For example, Palmeri et al. (1993) used a large set of 20 talkers to assess whether the recognition advantage observed for same-voice repetitions was attributable to the retention of gender information or to the retention of

more detailed voice characteristics. They hypothesized that if only gender information were retained, then there would be no expected differences in recognition between same-voice repetitions and different-voice/same-gender repetitions. Conversely, if more detailed information is retained, recognition deficits would be expected for words repeated in a different voice, regardless of gender. If only gender codes were prominent representations of spoken words, item recognition should be better for words repeated by talkers of the same gender than for words repeated by talkers of a different gender. Palmeri and colleagues found decrements in continuous recognition memory performance with different-voice repetitions, regardless of gender. Consistent with these findings, Church and Schacter (1992) and Schacter and Church (1994) found that voice change effects are mediated by F0, and not gender coding on implicit memory tasks. (Note, however, that these findings do not preclude the role of F0 as a cue for gender.)

Moreover, the role of F0 has been demonstrated in a variety of contexts. For example, Church and Schacter (1994) assessed the effects of changes in intonation and F0 with an implicit memory task. Intonation was manipulated by changing both emotional (happy or angry) and phrasal prosodic contour (question or statement) changes within a single voice. The manipulation of F0 consisted of raising and lowering the average F0 of a talker's voice by ten percent. Deficits in the priming performance on auditory identification and stem completion tasks following changes in F0 across study and test were observed. Moreover, all the manipulations that created voice change effects included significant changes in F0. These findings, along with those of Palmeri et al. (1993), provide evidence that pitch information plays a critical role in speech perception and spoken word recognition.

Speaking Rate Effects

Variability in speaking rate also has been shown to affect speech perception and spoken word recognition (Bradlow et al., 1999; Nygaard et al., 1995; Sommers et al., 1994). For example, Bradlow et al. (1999) manipulated rate by having talkers produce words at slow, medium, and fast rates. Sommers et al. (1994) showed a decrease in word identification scores for mixed speaking rate lists compared to single speaking rate lists. Other experiments that have utilized continuous recognition memory (Bradlow et al., 1999) and serial recall tasks (Nygaard et al., 1995) have obtained similar rate effects, such that poorer recognition and recall performance is obtained for multiple-rate lists relative to single-rate lists. In concert with Sommer et al. (1994), Bradlow et al. (1999) and Nygaard et al. (1995), the results indicate that variability introduced by changing speaking rate requires greater processing resources for encoding and rehearsal in working memory, thereby affecting the efficiency with which words are rehearsed and transferred into long-term memory. Like talker, as the speaking rate changes from trial to trial, fewer processing resources are available, which results in higher error rates and longer response times in high-variability rather than low-variability, contexts.

Amplitude and the Proposed Investigation

In contrast to the demonstrated effects of variability in speaking rate and F0, whether or not variability in the amplitude of an utterance affects phonetic processing still remains unclear. Overall amplitude, or the intensity at which a stimulus is presented, has not been found to affect phonetic or linguistic judgments of speech (Bradlow et al., 1999; Church & Schacter, 1994; Nygaard et al., 1995; Sommers et al., 1994). In fact,

researchers have failed to find deficits in identification scores or recognition memory performance for spoken word lists that had mixed overall amplitude relative to words presented at single overall amplitude (Bradlow et al., 1999; Church & Schacter, 1994; Nygaard et al., 1995; Sommers et al., 1994). This has led some researchers to conclude that amplitude is not a vocal parameter represented. For example, Bradlow and colleagues (1999) have suggested that the differential effects of talker and rate versus amplitude variability are due to fact that the talker and rate can have profound ramifications on the resolution of the spectral-temporal properties of segmental contrasts, whereas amplitude does not. Subsequently, Bradlow et al. (1999) contend that listeners might only be sensitive to variations in the speech signal that are phonetically relevant, that is, listeners devote a significant amount of processing resources only to perceiving and encoding changes in the speech signal that affect the nature and structure of their phonetic categorization and subsequent word recognition (Miller & Volaitis, 1989; Volaitis & Miller, 1992). Thus, Bradlow and her colleagues maintain that amplitude is not a phonetically relevant aspect of the speech signal.

However, there is evidence, which demonstrates that dynamic changes in amplitude are linguistically and phonetically relevant (Repp, 1982). This support stems from investigations that have examined the trading relation between voice onset time (VOT) and aspiration amplitude. VOT is the time interval between the consonantal noise burst that marks release of a stop closure and the onset of quasi-periodicity, which reflects laryngeal vibration. VOT is known to determine whether a particular stop consonant is perceived as voiced (e.g., /t/) or voiceless (e.g., /p/) regardless of the place of articulation (Lisker & Abramson, 1967). Aspiration is the noise generated by

turbulence as air moves through the glottis during the time the vocal folds are beginning to close for the following voiced sound. There exists a trade-off between VOT and aspiration amplitude, such that if the aspiration noise preceding the onset of voicing is increased, VOT can be decreased in order to maintain phonetic equivalence (Repp, 1982). This relationship could be argued to reflect well-documented, low-level effects of forward masking (e.g., Viemeister & Bacon, 1982), such that a high-intensity aspiration noise masks some of the voiced portion of the signal, resulting in perception that is more consistent with a voice-less phoneme. Alternatively, this trading relation could be argued to reflect processes specific to phoneme perception, as suggested by demonstrated influences of phonetic context on the perception of voicing contrasts (see Repp, 1982).

Additionally, research that has examined the role of vocal emotion in speech perception suggests that dynamic changes in amplitude appear to be linguistically relevant. Investigations of vocal paralanguage (i.e., non-word aspects of a voice such as intonation, pitch, sarcasm, and hesitations) have demonstrated that listeners use the vocal expression of emotion to extract the ultimate meaning of the utterance. For example, Reilly and Muzekari (1979) showed that in the face of discrepant propositional and vocal paralinguistic information, such as that found in sarcasm or joking, listeners consider the linguistic content (i.e., the words), but rely primarily on vocal paralanguage. Furthermore, the seminal work of Scherer (1986) suggests that there are specific acoustic features associated with the various emotions. For example, Scherer's (1986) descriptive findings indicate that speech produced in a joyful manner is consistently marked by increases in mean amplitude, whereas the converse is true of sadness. Subsequently, recent quantifiable data has demonstrated that listeners are sensitive to overall RMS

amplitude differences between joyful and sad word productions (Nygaard & Lunders, 2002). Although there is no direct support for the notion that dynamic changes in amplitude are important in speech perception, there is evidence that emotional cues are relevant to extracting the true meaning of an utterance. It is possible that one way of encoding emotional information is through the specification of an amplitude contour. This contour might be stored where it can later be accessed, to facilitate spoken word recognition.

CHAPTER NOTE

[1]Representation refers to information that can be retrieved in the context of a given task. Given the focus of the current investigation on the nature of information that is represented, including whether word and voice information is represented jointly or separately, the data do not permit claims about the neurological processes that realize mental representations. It is acknowledged that there is much debate among psychologists and philosophers of the mind surrounding issues of representation. At the center of this debate are the following major issues: 1) the ontological status of the psychological in relation to the physical and 2) what is or is likely to be the best current future theory of cognitive processing (Greenwood, 1991).

CHAPTER 3

PRESENT STUDY OVERVIEW

Experiment 1

The proposed investigation sought foremost to determine if dynamic changes in the amplitude of a speech waveform are represented along with word information. An emotional manipulation was used to examine if listeners are sensitive to dynamic changes in amplitude. Because there exists little quantitative information about the vocal aspects associated with emotion, Experiment 1 was a word production study designed to verify whether there are specific changes in amplitude associated with the production of a given emotion. The purpose of Experiment 1 was to investigate whether words produced with the same intended emotional valence share a similar amplitude contour and whether they differ across emotions. Another goal was to create a set of stimuli that could be used in the recognition memory task of Experiment 2. Specific amplitude measurements were obtained and compared for each production.

Based on previous findings (Nygaard, 2002; Scherer, 1986), it was expected that there would be an overall difference between emotions with respect to both overall amplitude and dynamic changes in amplitude, namely that joyful productions would reflect greater levels of amplitude relative to sad productions. The words produced in Experiment 1 were used to generate the stimuli in Experiment 2.

Experiment 2

Experiment 2 sought to determine whether the dynamic variability in amplitude influences recognition memory for spoken words. Talker and amplitude variability were manipulated by imposing artificial amplitude contours associated with joy and sadness on emotionally neutral concatenated speech productions. The representation of amplitude information along with word information would be indicated by improved recognition memory performance when items are repeated in the same amplitude contour.

A secondary purpose of Experiment 2 was to replicate previous demonstrations of talker effects. In this way, if the amplitude contour manipulation failed to affect recognition memory performance, replication of talker effects would serve to ensure that the lack of amplitude contour effects are not due to the nature of the stimuli, but rather, to the amplitude contour manipulation.

Experiment 3

Experiment 3 was a perceptual discrimination study that assessed whether listeners were sensitive to amplitude contour information, that is, we examined whether listeners could reliably perceive differences between joyful and sad amplitude contours imposed on consonant vowel (CV) syllables. Moreover, this experiment sought to determine whether the amplitude manipulation used in Experiment 2 was perceptually salient. If listeners' can discern differences between joyful and sad amplitude contours, then this would suggest that listeners are capable of using amplitude contour information during speech perception, and thus, could potentially represent amplitude contour information.

CHAPTER 4

EXPERIMENT 1: METHODOLOGY AND RESULTS

Participants

Participants included six students (3 females and 3 males) enrolled in psychology courses at the University of Nevada, Las Vegas. Four of the participants were undergraduate assistants in the UNLV Auditory Perception Laboratory. The other two participants consisted of a psychology graduate student and a part-time undergraduate student. All were native speakers of American English who reported no history of speech or hearing disorders.

Stimuli

The stimuli used in Experiment 1 were 200 monosyllabic words drawn from four different 50-item phonetically balanced (PB) word lists (ANSI, 1971). PB word lists approximate the relative frequency of a phoneme's occurrence in a language and have been traditionally used in word recognition tests to measure speech perception abilities (Pisoni, Miyamoto, Kirk, Sommers, & Osberger, 1994). Subsequently, PB word lists have been useful because they eliminate phoneme-based interactions, and thereby help to ensure that the obtained effects are due to talker. The PB word lists that were used were

identical to those used in other investigations of talker variability (see Bradlow et al., 1999; Nygaard et al., 1994, 1995; Nygaard & Pisoni, 1998; Sommers et al., 1994).

Like the selection of the PB word lists, the carrier phrase "Please say the word ___" was adopted from similar talker variability investigations (see Bradlow et al., 1999; Nygaard et al., 1994, 1995). The use of the carrier phrase was necessary to reduce the effects on contour that are the result of producing isolated words in a list. Each word from the lists was embedded in the carrier phrase to ensure that each word was produced in the same sentence context.

All productions were recorded in an Acoustic Systems single-walled sound attenuated chamber. The productions were transduced with a Sennheiser MD735 microphone and digitized at a sampling rate of 16 kHz (16-bit stereo). Productions for a word list in a given emotion were recorded as a continuous file and later separated into individual word files. Syntrillium Corporation's Cool Edit Pro (1999) wave editing software was used to remove each word production from the carrier phrase.

Procedure

Participants were seated in the sound-attenuated chamber and produced into a microphone each word embedded in the carrier phrase from four typed versions of different 50-item PB word lists. Each list was produced in an emotionally-neutral manner, as well as in a manner that was associated with each of the following emotions: joy and sadness. These emotions were selected based on Plutchik's (1980) eight-factor "wheel" model of emotions, which includes joy and sadness as two of the eight primary emotions. Accordingly, joy and sadness are arranged as opposing pairs of emotions in the

model. The selection of such opposing emotions in the proposed investigation was motivated by the notion that such emotions would be more likely to reflect acoustic differences in production relative to two adjacent emotions on Plutchik's wheel, such as grief and sadness. The more similar emotions are to each other, the more likely that they could be produced utilizing the same vocal manner. Although other models of emotion (e.g., Arnold, 1960; Ekman, Friesen, & Ellsworth, 1972; Izard, 1971) define the primary emotions somewhat differently, the target emotions (i.e., joy and sadness) are generally included in these alternative models. Moreover, common alternative models share with Plutchik's wheel the notion that these emotions are universal in that they appear to exist in all cultures, and some can be identified in higher animals (Ortony & Turner, 1990).

Following the completion of a word list with or without an intended affect, participants were provided three- to five-minute rest breaks to prevent fatigue. Participants were given an unlimited amount of time to produce each word list and were permitted to make multiple attempts at production. Only the last production of each word was used because it presumably reflected the talker's most accurate production of the intended affect. During the recording session, an experimenter monitored participants' performance from a computer located outside of the booth to ensure that each word was pronounced correctly. Both emotion and word lists were counterbalanced to prevent potential order effects. Recordings were made in a single session that lasted an average of one hour and thirty minutes for each participant.

Amplitude Measurements

Following the editing process, time recorded in milliseconds (msec) and relative amplitude (in dB; 0 = maximum value) was collected at 11 points within the final productions of each word at each emotion. Measurements were obtained at the beginning and end of the word, as well as at increments that corresponded to every ten percent of the production's duration. More specifically, measurements were taken at the highest closest occurring positive peak amplitude within the cycle of the waveform. One of the characteristics of voiced speech is periodicity, that is, there are repeating patterns across the signal. A cycle represents one occurrence of this pattern. If a point of measurement falls somewhere within the cycle, the value was always collected at the peak of the next cycle in the waveform.

Design and Analysis

An items analysis was conducted in the form of a 6 (Talker) x 2 (Gender) x 2 (Affect) x 11 (Time) mixed-model Analysis of Variance (ANOVA). Affect and Time as the within-subjects variables and Talker and Gender as the between-subjects variables was conducted to determine if there were overall differences in the amplitude contours (i.e., as measured in dB) between affect and talkers. The two levels of the factor Affect were joy and sadness (amplitude contours). The two levels of the factor Gender were male and female. The factor Talker had six levels; each represented a different talker (3 females and 3 males). The factor Time consisted of 11 levels, of which two of the levels represented signal onset and offset, whereas the other 9 levels represented increments that

corresponded to 10% of the words' total duration. Amplitude (in dB) was measured at each of these levels of time.

Results

There were distinct differences in the manner in which talkers produced items from the word lists. Figure 1 displays the mean amplitude contours for joyful and sad productions for each talker. As can be seen in the figure, there were some general differences between the productions of talkers'. Evidence for overall differences in production was supported by ANOVA results, which revealed a statistically significant main effect for Talker, $F(1, 294) = 14.932, p < 0.01$. Scheffe's post hoc tests indicated that talker 6 produced lower amplitude utterances relative to all other talkers', $p < 0.05$. This main effect of talker could be due to the fact that the talkers with lower amplitude productions could have a general tendency to speak more softly than others. Alternatively, this main effect could simply reflect the fact that talkers who produced low amplitude contours were at a greater distance from the microphone during recording. Thus, the variability in overall amplitude in the talker set should not be unexpected, but rather should be considered a typical pattern of variability across talkers and/or recordings.

Although talker differences were observed, there was not a statistically significant main effect of Gender, $F < 1$, nor were there any significant interactions involving that variable. This null finding indicates that there were not significant differences in the manner in which males and females produced words.

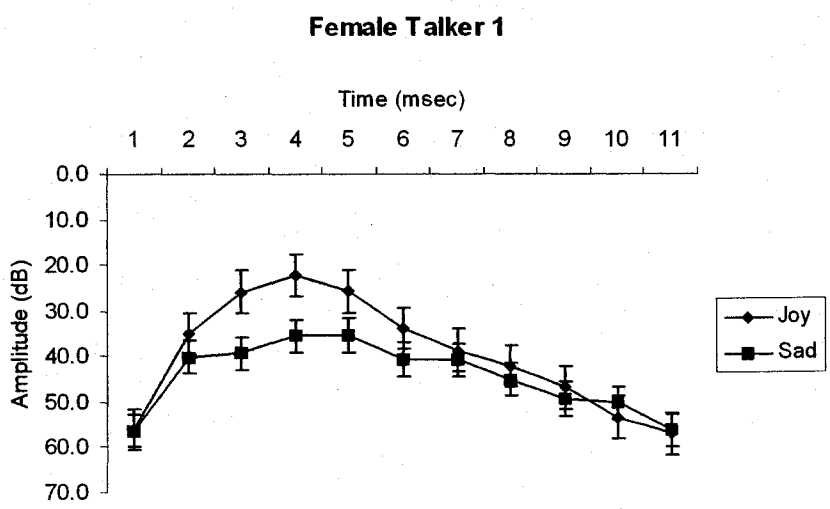

Figure 1. Mean amplitude contours for joy and sad productions obtained at 11 time intervals (corresponding to 10% increments) of each production for female talker 1.

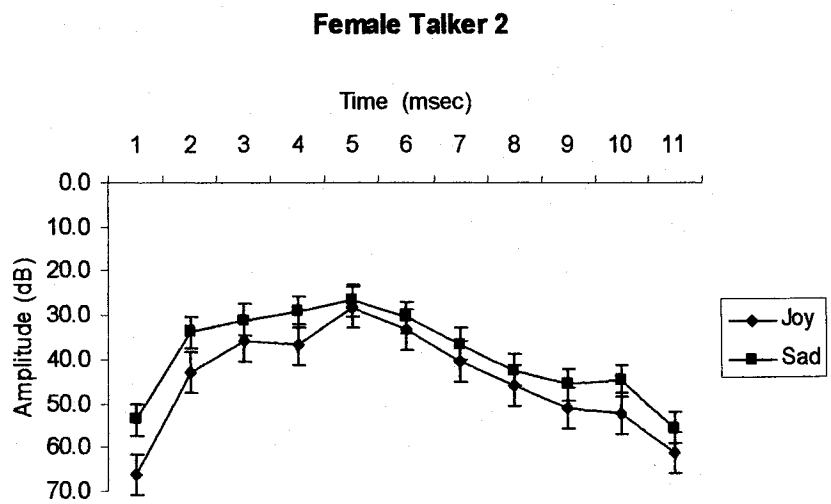

Figure 2. Mean amplitude contours for joy and sad productions obtained at 11 time intervals (corresponding to 10% increments) of each production for female talker 2.

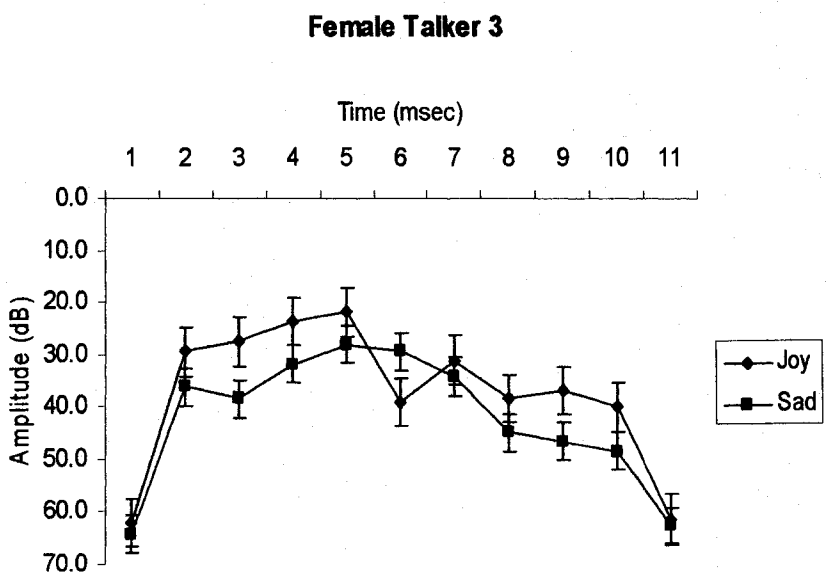

Figure 3. Mean amplitude contours for joy and sad productions obtained at 11 time intervals (corresponding to 10% increments) of each production for female talker 3.

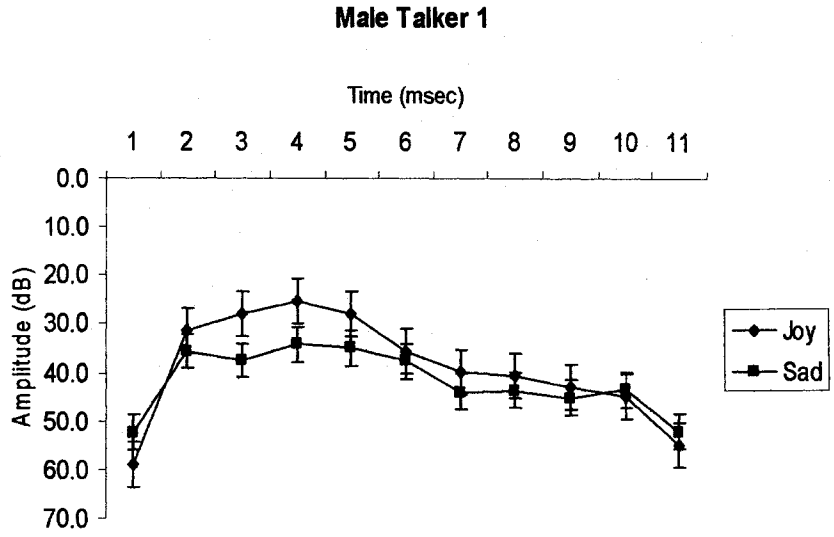

Figure 4. Mean amplitude contours for joy and sad productions obtained at 11 time intervals (corresponding to 10% increments) of each production for male talker 1.

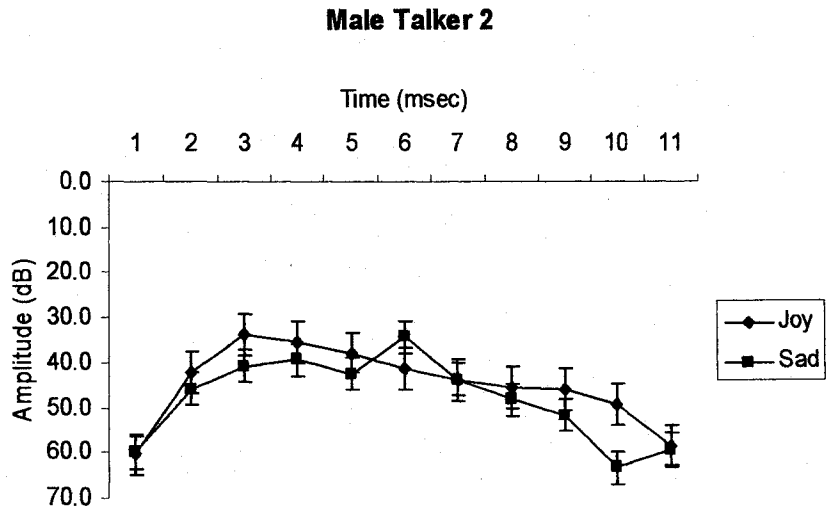

Figure 5. Mean amplitude contours for joy and sad productions obtained at 11 time intervals (corresponding to 10% increments) of each production for male talker 2.

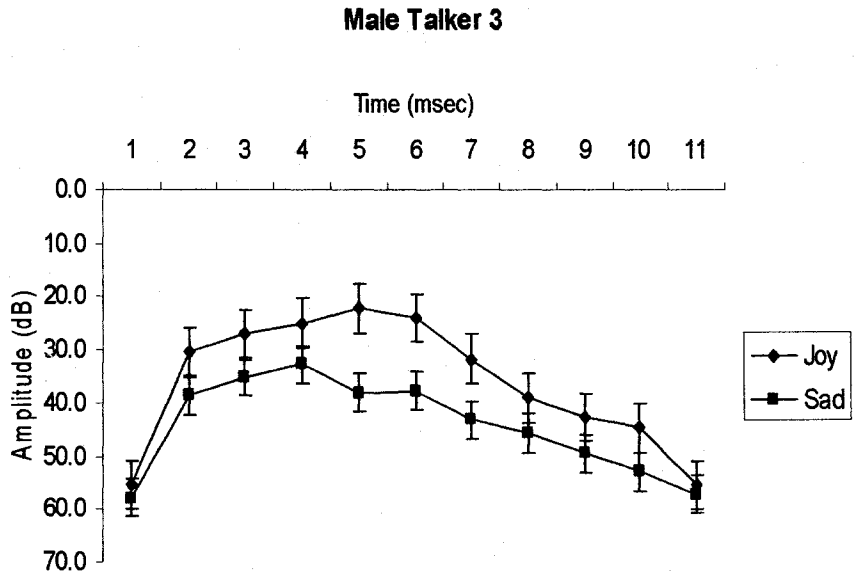

Figure 6. Mean amplitude contours for joy and sad productions obtained at 11 time intervals (corresponding to 10% increments) of each production for male talker 3.

There were some systematic differences in amplitude across the 11 time intervals, which was demonstrated by a main effect for Time, $F(1, 294) = 367.528$, $p < 0.001$. Because the sphericity assumption was not met, Mauchly's $W(54) = 0.064$, $p < 0.01$, the

Huynh-Feldt correction was applied to the F value derived for Time. Post-hoc pair-wise comparisons were made using the Fisher-Haytar (1986) range test. Amplitude increased from signal onset, with a significantly sharper increase in the middle of the signal [time 6, $M = 30.312$ *(0.501)*], followed by a corresponding significant decrease in amplitude at offset [time11, $M = 48.256$ *(0.427)*], $q(9, 294) = 26.148$, $p < 0.01$. The fact that amplitude levels are lowest at signal onset and offset is not surprising because it is a necessary consequence of speech production. After all, onset represents the initiation of airflow through the vocal tract, and offset reflects the closure of airflow. Also, amplitude at signal offset (time 11, $M = 48.256$ *(0.427)*) was significantly greater than at signal onset (time 1, $M = 58.631$ *(0.796)*), $q(3, 294) = 15.118$, $p < 0.01$. This finding may be a consequence of the fact that each word was produced in a sentence-final position in the context of a list of sentences. For example, talkers had slightly higher levels of amplitude at offset in anticipation of the beginning of the next sentence production.

A Talker x Time interaction, $F(5, 294) = 3.834$, $p < 0.01$ also was observed, indicating that the amplitude contours of individual talkers varied substantially across talkers around the middle half of the utterance. This was confirmed by tests of simple effects, which revealed that talkers' productions of amplitude differed significantly at time 3, $F(5, 294) = 3.109$, $p < 0.01$, time 5, $F(5, 294) = 2.196$, $p < 0.05$, time 6, $F(5, 294) = 3.192$, $p < 0.01$, time 7, $F(5, 294) = 3.009$, $p < 0.01$, and time 8, $F(5, 294) = 2.666$, $p < 0.02$, respectively. These effects appear to be due to differences across talkers with respect to the magnitude of increases or decreases in amplitude as a function of time, as well as the time at which peak amplitude occurred. For an example, compare the average contours of male talker 2 and female talker 3, in Figure 1. Relative to this female talker,

male talker 2 produced a peak amplitude that was more than 12dB lower, and that occurred much earlier in the signal (time 4 versus time 6 for the female).

The experiment was primarily interested in examining the effects of intended emotion on the amplitude of talkers' productions. Several effects of emotion were observed. For example, a main effect for Affect, $F(1, 294) = 98.689$, $p < 0.01$ was revealed, such that the mean intensity was more than doubled (just over 3 dB) for joy productions compared with sad productions. Thus, it appears that talkers generally increase amplitude when they intend to convey the emotion joy.

Individual differences also were observed in the production of amplitude as a function of affect, resulting in an Affect x Talker interaction, $F(1, 294) = 38.644$, $p < 0.01$. (Degrees of freedom were adjusted using the Satterthwaite (1946) solution). Closer examination of the data with simple effects reveals the basis for this interaction. Mean amplitude did not change across talkers for sadness, $F(1, 294) = 1.271$, $p > 0.27$. In contrast, mean amplitude for joy changed significantly across talkers, $F(1, 294) = 2.684$, $p < 0.05$. Specifically, mean amplitude was greater for joyful productions relative to sadness for two talkers [male talker 3, $F(1, 294) = 10.905$, $p < 0.01$; female talker 3, $F(1, 294) = 5.821$, $p < 0 .05$] and marginally greater for a third talker [female talker 1, $F(1, 294) = 3.554$, $p > 0.05$]. In contrast, two talkers produced no difference in mean amplitudes for joy and sadness [male talker 1, $F(1, 294) = 1.215$, $p > 0.05$; male talker 2, $F < 1$]. Finally, female talker 2 produced significantly higher mean amplitudes for sadness [$F(1, 294) = 10.905$, $p < 0.01$; see Figure 1]. Thus, while there are substantial individual differences in the mean amplitudes of productions they appear to be restricted to a single emotion—joy.

More importantly, talkers systematically produced differences in joyful and sad amplitude contours, as indicated by an Affect x Time interaction, $F(1,294) = 12.033$, $p < 0.01$. Tests of simple effects confirmed that there were larger fluctuations in amplitude across productions of joy relative to sad, $F(1, 294) = 39.044$, $p < 0.01$. Specifically, joyful contours tended to be higher in amplitude during the middle portion of the signal [time 4, $F(1, 294) = 8.047$, $p < 0.01$; time 5, $F(1, 294) = 6.768$, $p < 0.01$; time 6, $F(1, 294) = 5.465$, $p < 0.02$, respectively]. In contrast, examination of amplitude at the early portion of the signal [time 1, $F(1, 294) = 15.118$, $p < 0.001$) and at the latter portion of the signal (time 11, $F(1, 294) = 48.256$, $p < 0.001$) did not differ significantly between joyful and sad contours, $p > 0.10$. A good example of these general tendencies can be seen in the mean amplitude contours for joyful and sad productions by female talker 1 in Figure 1. Given that talkers actually produce these dynamic (time-based) differences with emotion, it is likely that this could represent a perceptual cue to intended affect. The observed consistent differences between joy and sad amplitude contours provided an appropriate manipulation to be used in Experiment 2.

An overall trend analysis was also conducted to evaluate the basic shape of the amplitude contours associated with each emotion. The results of this analysis are summarized in Table 1. As indicated by the asterisks in Table 1, significant linear, quadratic, and cubic trends were found for productions of words intended to convey joy, whereas significant quadratic and cubic trends were observed for sad words. The significant quadratic trends illustrate the general curvilinear nature of both joyful and sad amplitude contours. The curvilinear shape corresponds to an initial rise in amplitude from signal onset, to a plateau in the middle portion of the signal, followed by a gradual

decline in amplitude at signal offset. At least some of this trend should be taken to reflect basic constraints of speech production, namely zero amplitude at signal onset and offset, and a peak amplitude being reached at some time during the utterance. In addition, joyful productions also exhibited a cubic trend. This trend likely reflects talkers' tendency to produce higher amplitudes at signal offset compared to signal onset, presumably due to the aforementioned anticipation of the subsequent utterance.

Table 1 Trend analysis (computed across all talkers) summary table for each emotion

			Joy	
Method	Rsq	DF	F	Significance
Linear	0.043	298	13.45	0.001*
Quadratic	0.060	297	9.55	0.001*
Cubic	0.254	296	14.61	0.001*
			Sad	
Method	Rsq	DF	F	Significance
Linear	0.010	298	3.00	0.084
Quadratic	0.129	297	22.03	0.001*
Cubic	0.130	296	14.68	0.001*

Finally, the relationship between joyful and sad amplitude contours changed as a function of talker. This change was reflected in a three-way interaction of Time x Affect x Talker, $F(5, 294) = 2.494$, $p < 0.01$. A trend analysis was conducted to evaluate the shape of the relationship between time and amplitude for each talker and across affect. Significant linear, quadratic, and cubic trends were found for joy and sadness for most talkers (see Table 2). Thus, with a few exceptions (e.g., Talker 2), it appears that the shapes of the contours for each emotion determined from the overall trend analyses above do reflect general production tendencies across (at least this subset of) talkers. Of course, some individual differences in production were observed, such that the productions of

some talkers [e.g., Talker 1 in Table 2 (Female Talker 1 in Figure 1)] tended to better approximate a systematic curvilinear rise and fall of amplitude than others [e.g., Talker 5 in Table 2 (Male Talker 2 in Figure 5)]. These individual differences can be clearly seen by comparing average amplitude contours across talkers in Figure1.

Table 2 Trend analysis summary table for each individual talker and across each emotion

				Joy	
Talker	Method	Rsq	DF	F	Significance
1	Linear	0.147	48	8.25	0.006*
	Quadratic	0.230	47	7.02	0.002*
	Cubic	0.254	46	5.23	0.003*
2	Linear	0.011	48	0.56	0.459
	Quadratic	0.012	47	0.29	0.747
	Cubic	0.124	46	2.18	0.103
3	Linear	0.119	48	6.47	0.014*
	Quadratic	0.199	47	5.85	0.005*
	Cubic	0.199	46	3.82	0.016*
4	Linear	0.055	48	2.77	0.102
	Quadratic	0.198	47	5.79	0.006*
	Cubic	0.212	46	4.13	0.011*
5	Linear	0.066	48	3.40	0.072
	Quadratic	0.068	47	1.73	0.189
	Cubic	0.069	46	1.13	0.346
6	Linear	0.106	48	5.67	0.021*
	Quadratic	0.130	47	3.52	0.038*
	Cubic	0.130	46	2.30	0.090
				Sad	
Talker	Method	Rsq	DF	F	Significance
1	Linear	0.088	48	4.65	0.036*
	Quadratic	0.240	47	7.44	0.002*
	Cubic	0.251	46	5.12	0.004*
2	Linear	0.023	48	1.12	0.296
	Quadratic	0.078	47	1.99	0.148
	Cubic	0.272	46	5.74	0.002*
3	Linear	0.112	48	6.06	0.017*
	Quadratic	0.127	47	3.43	0.041*
	Cubic	0.141	46	2.51	0.070
4	Linear	0.001	48	0.02	0.877
	Quadratic	0.034	47	0.83	0.444
	Cubic	0.036	46	0.58	0.633
5	Linear	0.140	48	7.84	0.007*
	Quadratic	0.191	47	5.56	0.007*
	Cubic	0.193	46	3.67	0.019*
6	Linear	0.215	48	13.16	0.001*
	Quadratic	0.237	47	7.29	0.002*
	Cubic	0.239	46	4.81	0.005*

In summary, talkers exhibited idiosyncratic differences in their production of words intending to convey joy and sadness. The magnitude of increase or decrease in amplitude across time, as well as the time at which peak amplitude occurred varied

among talkers'. The greatest individual differences in the amplitude contours of talkers' productions occurred for joy. In contrast, talkers showed general trends or consistencies in production, such as the tendency to increase amplitude when the intent was to convey joy. Also, the mean amplitude did not change across talkers for sadness.

The observed consistent differences between joy and sad amplitude contours provided an appropriate manipulation to be used in Experiment 2. Given that talkers produce these dynamic (time-based) differences with emotion, it is likely that this could represent a perceptual cue to intended affect.

CHAPTER 5

EXPERIMENT 2: METHODOLOGY AND RESULTS

Overview

This experiment sought to determine if amplitude contours (i.e., dynamic, as opposed to static, manipulations) are encoded and represented with word information. The spoken word recognition paradigm employed by Bradlow et al. (1999) was utilized in this experiment. A secondary goal of this experiment was to replicate talker effects. Talker effects have been shown to be a robust phenomenon so far as they have been frequently replicated. As a result, the talker manipulation served to ensure that participants were trying to perform the task.

Based on previous research (Goldinger et al., 1991; Mullenix et al., 1989; Nygaard et al., 1995; Palmeri et al., 1993), talker variability was expected to affect recognition memory for spoken words. Items repeated by the same talker were expected to be better recognized than items repeated by a different talker. Furthermore, consistent with Bradlow et al.'s (1999) findings, accuracy was predicted to decrease with increasing lag, for both same and different talker lists.

Similarly, the variability introduced by changes in amplitude contour was expected to affect recognition memory for spoken words. Words repeated in the same amplitude contour were predicted to be recognized more accurately than words repeated in a different amplitude contour. This would suggest that variations in amplitude contour

also require processing (e.g., compensation, normalization), which compete for a limited pool of resources. Finally, accuracy was expected to decrease with increasing lag for both same and different amplitude contour conditions.

Participants

Twenty students enrolled in introductory psychology courses at the University of Nevada, Las Vegas participated in this experiment. All participants were native speakers of American English who reported no history of speech or hearing disorders at the time of testing. All participants received partial course credit for their participation.

Stimuli

The stimuli used in Experiment 2 consisted of the PB words used in Experiment 1. These words were input into AT&T's Natural Voices Text-to-Speech engine at a 16 kHz (16-bit) sample rate. Text-to-speech engines convert written text to speech output through a concatenation approach. The concatenation method generated synthetic speech by splicing together segments (i.e., phonemes) of encoded speech. Concatenated speech incorporates contextual information, that is, it accounts for coarticulation (the coproduction or overlapping of neighboring speech sounds). A mathematical algorithm then employed a smoothing function, which acts to make productions more human-like. This method of stimulus generation was employed for two purposes. First, it minimized the degree of affective information conveyed in the productions, which could threaten the validity of the manipulation. Secondly, it provided a relatively noise free recording. Stimuli were generated for both a male and female talker of American English.

Amplitude Contours

Artificial contours associated with joy and sadness were imposed over the concatenated speech productions. Given that talkers demonstrated similar trends in their productions of joy and sad amplitude contours (Experiment 1; see figure 2), amplitude contours for the talker that reflected maximal differences between joy and sadness were selected. Syntrillum's Cool Edit Pro (1999) was used to impose the amplitude contours on each word production. Amplitude envelopes were altered linearly from one measured percentage point to the next in order to match the extreme contours.

Amplitude levels for all stimuli were normalized and equated for RMS to minimize overall loudness differences. Amplitude was then ramped linearly over the final 20 ms of each production.

Stimuli were presented binaurally at a peak intensity of 80 dB[A] over Sennheiser HD25 headphones. At presentation, the stimuli were submitted to a low-pass Butterworth anti-aliasing filter with a cut-off frequency of 4.8 kHz (-24 dB/octave skirt). All testing took place in an Acoustic Systems RE-142 single-walled sound attenuated chamber. A PC presented the stimuli and collected participant responses via Kendall's (2000) *Music Experiment Development System (MEDS 2001-A)*.

Word Lists

Two word lists were constructed to resemble those of Bradlow et al. (1999), in which each test word was presented and then repeated once after a lag of 2, 8, 16, or 32 intervening items. The test word itself counted as the first intervening item. Lag was manipulated to assess the effects of both talker and amplitude on perceptual coding and

rehearsal. Previous research has consistently observed strong interactions between the length of the lag and talker and rate variability (Bradlow et al., 1999; Goldinger et al., 1991; Nygaard et al., 1995; Palmeri et al., 1993; Sommers et al., 1994). Specifically, accuracy has been found to decrease with increasing lag. Moreover, greater lag affords less processing time and interferes with efficient rehearsal processes, which result in decrements in both identification and memory performance. Lag between the first and second repetition of a word was manipulated as a within-subjects variable (2, 8, 16, or 32 words), as in Bradlow et al.'s (1999) manipulation.

Each list began with 15 practice words, which were used to familiarize participants with the test procedure. Practice trials consisted of eight words, seven of which were repeated and the other acted as a filler item. The practice words were spread across the four different conditions and none of these 15 words was repeated in the experiment. The next 30 trials were used to establish a memory load and were not included in the final data analyses. The memory load items were non-repeating. A memory load was used to equate performance on stimulus pairs occurring early in the list with pairs occurring later in the list. Discarding the first 30 trials permitted the evaluation of memory performance in participants whose memory buffers were already full (Bradlow et al., 1999).

The actual test list consisted of 144 test word pairs. Twenty-one filler items were interspersed in the test list and were not included in the analyses. There were nine test words per combination of talker condition with each lag. The total number of items in each list was 354. Two separate randomizations of the lists were created. Participants

were randomly assigned to the two lists. For the final analyses, data were collapsed across randomizations.

Conditions

Talker (male and female) and amplitude contour (joy and sadness) for the second repetition of test items were manipulated to produce four conditions. The conditions were as follows: talker-same-amplitude-same (TSAS), talker-same-amplitude-different (TSAD), talker-different-amplitude-same (TDAS), and talker-different-amplitude-different (TDAD). In the TSAS condition, both the talker (male or female) and the amplitude contour (joy or sadness) remained the same between the first and second repetition of the word. In the TSAD condition, the talker remained the same, but the amplitude contour changed between the first and second repetition. In the TDAS condition, talker changed across repetitions, but the amplitude contour remained the same. Finally, in the TDAD condition, both the talker and the amplitude contour changes from the first to the second repetition. Condition was manipulated as a within-subjects variable. In each condition, there were nine words per lag and a mathematical equation was employed to equate for emotion.

Procedure

The word recognition task was similar to the one used by Bradlow et al. (1999). On each trial, listeners heard a spoken word and were asked to identify whether the word was "new" (i.e., the word was new to the list) or "old" (i.e., the word had appeared previously in the list of spoken words). Listeners were instructed to make their responses

as quickly and accurately as possible via key presses on a computer keyboard. The "1" key corresponded to a judgment of "new" and the "3" key reflected a judgment of "old". Listeners had an unlimited amount of time to enter their responses and no feedback was provided. The entire session of 354 trials lasted approximately 25-30 minutes.

Design and Analysis

A 2 (Talker) x 2 (Amplitude) x 2 (First Emotion) x 4 (Lag) repeated measures ANOVA was conducted. The two levels of Talker and Amplitude were same and different and the two levels for the factor First Emotion were joy and sadness. The factor First Emotion referred to the emotion that was presented on the first repetition of a word. The four levels of lag were 2, 8, 16, and 32. The dependent measure of interest was accuracy (overall percent correct).

Results and Discussion

There was an initial concern that recognition might be affected by which emotion contour accompanied the first presentation of a word. This concern arose from reports of lab personnel during stimulus preparation that words with the sad contour tended to seem longer in duration, whereas words with the joy contour seemed noticeably briefer. This perception existed despite the fact that word duration was identical across both types of contours. One possible explanation for this anecdotal report comes from the relative shape of the contours. As demonstrated in Experiment 1 (for an example with talker 1, see Figure 1), joyful amplitude contours reflected higher peak amplitudes than sad contours at the middle portion of the signal. One consequence of this elevation in peak

amplitude for the joyful stimuli was a sharper decrease in amplitude immediately following the peak. It is likely that such a sharp decrease in amplitude conveys to the listener that the stimulus is rapidly coming to an end. Despite these concerns, recognition accuracy was not influenced by which emotion (or amplitude contour; joy versus sad) appeared in the first presentation of an item (i.e., at study). This was indicated by the failure to find either a statistically significant main effect of First Emotion, $F < 1$, or any interactions with that variable.

Word recognition performance in Experiment 2 was affected by variables that have been previously demonstrated to influence recognition memory for words. For example, a main effect for Lag was obtained, $F(3, 57) = 21.087$, $p < 0.01$. Apriori Bonferroni comparisons showed that accuracy decreased with increasing lag, $F(1, 57) = 7.231$, $p < 0.01$, such that accuracy was greatest at the shorter lags (i.e., 2 and 8) and lowest at the larger lags (i.e., 16 and 32). Mean accuracy (i.e., percentage correct) at each lag with the accompanying standard errors are as follows: lag 2, $M = 93$ *(0.21)*, lag 8, $M = 85$ *(0.18)*, lag 16, $M = 81$ *(0.36)*, and lag 32, $M = 77$ *(0.33)*. Thus, it is clear that the procedure produced a memory load as lag increased. Such an effect of lag is robust and has been well established in the word recognition literature (e.g., Bradlow et al., 1999; Church & Schacter, 1994; Goldinger et al., 1991; Nygaard et al., 1995; Palmeri et al., 1993; Schacter & Church, 1992; Sommers et al., 1994). Potential interactions between lag and other variables failed to reach significance in analyses.

The anticipated effects of talker on recognition memory performance also were obtained. Recognition accuracy was significantly better for test pairs that shared the same talker (same talker conditions, $M = 87\%$; different talker conditions, $M = 79\%$), as

indicated by a main effect for Talker, $F(1, 19) = 43.289$, $p < 0.001$. Figure 2 shows the word recognition accuracies (overall mean percent correct "old" responses) for each talker and amplitude condition and collapsed across lag. The obtained effect of talker can be seen when the means in Figure 2 are collapsed across amplitude conditions. This result replicates previous findings (Bradlow et al., 1999, Experiment 1; Church & Schacter, 1994; Goldinger et al., 1991; Nygaard et al., 1995; Palmeri et al., 1993; Schacter & Church, 1992; Sommers et al., 1994), which showed a same talker advantage for recognizing a word as a repeated item without any explicit instructions to the listeners to attend to the identity of the talker. In addition, separate ANOVA results examining talker gender indicated there was no difference in recognition accuracy for words spoken by a female talker versus a male talker, $F(1, 19) = .074$, $p > 0.05$. The absence of a gender advantage contrasts with Bradlow et al.'s (1999) finding of a slight advantage for the male talker. These data, however, complement the finding of a general effect of talker. The talker effects observed are consistent with the representation of talker information, as previously argued (Bradlow et al., 1999, Experiment 1; Church & Schacter, 1994; Goldinger et al., 1991; Nygaard et al., 1995; Palmeri et al., 1993; Schacter & Church, 1992; Sommers et al., 1994).

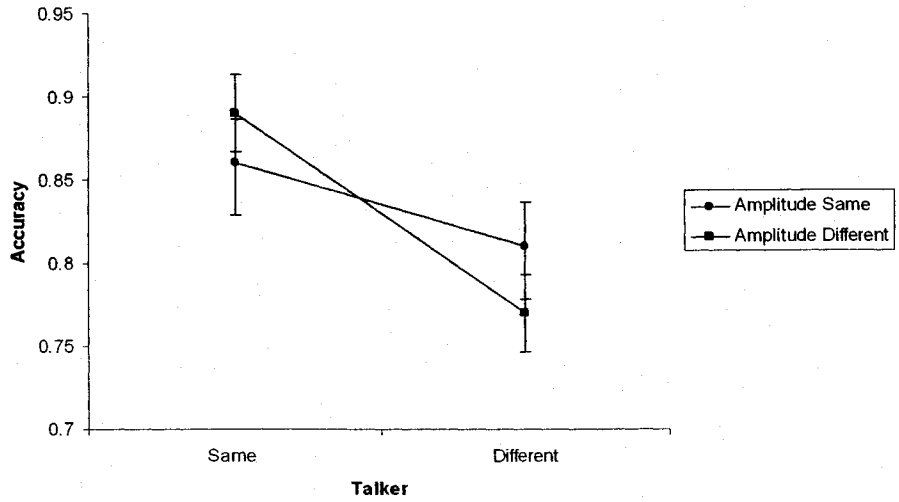

Figure 7. Overall percentage correct old word recognition responses (with standard errors) for each talker condition and collapsed across lag.

If amplitude (contour) impacts recognition performance in a similar manner as talker, then one should expect to see recognition accuracy increase when amplitude is consistent across repetitions of a given word item. The additional potential benefit to recognition performance by both amplitude and talker information remaining the same across repetitions of the word could be either additive or multiplicative in nature. Some aspects of an observed Talker x Amplitude interaction, $F(1, 19) = 4.88$, $p < 0.05$, were initially consistent with a potential weak influence of amplitude on word recognition performance. Specifically, in the conditions involving different talkers (i.e., TDAS, $M = 84\%$; TDAD, $M = 80\%$), performance was slightly (3 percent) better when amplitude remained the same than when it changed (see Figure 2). Interestingly, an opposing (4 percent) tendency was obtained in the same talker conditions (i.e., TSAS, $M = 89\%$; TSAD, $M = 91\%$). However, neither of these tendencies for amplitude was statistically significant. In contrast, post-hoc pair-wise comparisons (Student-Neumann

Kuels range test) showed that talker effects largely influenced recognition accuracy. Accuracy was always better in conditions where the talker remained the same compared to conditions where the talker changed [TSAD versus TDAD, $q(4, 19) = 5.771, p < 0.001$; TSAS versus TDAD, $q(3, 19) = 4.553, p < 0.01$; TSAD versus TDAS, $q(3, 19) = 3.812, p < 0.02$]. Taken together these findings suggest that changes in talker affect recognition accuracy, whereas changes in amplitude do not.

Given that the dynamic manipulation of amplitude did not impact word recognition accuracy in a consistent or predictable manner, it could be argued that amplitude contour information may not be a source of variability in speech that is encoded into long-term memory in the same way as talker and rate information. Specifically, dynamic information about amplitude may not be represented directly with its associated word(s). This possibility for a time-varying manipulation of amplitude is consistent with the conclusions of Bradlow, et al. (1999) for static manipulations of (overall) amplitude.

Alternatively, it could be argued that the amplitude manipulation may have been too subtle to affect performance. If so, it could still be plausible for amplitude contour information to be represented with other word information. According to this alternative argument, the absence of predictable amplitude effects on recognition performance must be due to insufficient variability in amplitude across the two types of contours. As a result, differences in contour could not be perceptually registered, and thus could not be encoded. Fortunately, this latter alternative can be directly addressed. This constituted the goal of Experiment 3.

CHAPTER 6

EXPERIMENT 3: METHODOLOGY AND RESULTS

Overview

Experiment 3 was conducted to insure that listeners could reliably perceive the amplitude contours used in Experiment 2. This was accomplished by examining listeners' perceptual ability to discriminate the two types of amplitude contours (associated with joyful and sad affect). Listeners were given two consonant vowel (CV) syllables on each trial and were required to identify whether the amplitude contour imposed on the CV in the first interval relative to the CV in the second interval corresponded to the emotion of joy or sadness. It was hypothesized that if listeners are sensitive to the emotional information that the contours convey, then they should be able to accurately identify and label the emotion associated with CVs that have an artificially imposed amplitude contour. The demonstration that listeners can reliably perceive emotional differences between the contours associated with joy and sadness would suggest that the dynamic amplitude manipulation was perceptually salient. Furthermore, reliable discrimination performance would also suggest that dynamic changes in amplitude might be retained to some extent.

Participants

Ten listeners served as subjects. One of the listeners also produced the stimuli used in this experiment. Given that this subject only produced the syllables and not the final contours, this subject's inclusion was not a concern. Participants' only task in this experiment was to distinguish joy from sad contours, not the syllables themselves. All participants were graduate and undergraduate lab assistants in the UNLV Auditory Perception Laboratory, with the exception of one participant who was a graduate student from another UNLV laboratory. Some of the listeners had previous experience with the amplitude contours, in that they were involved in imposing the amplitude contours on PB words used in Experiment 2. Each participant was a native speaker of American English with no history of speech or hearing disorders.

Stimuli

The stimuli used in Experiment 3 were CV syllables. Consonants varied in terms of their place of articulation (bilabial, alveolar, and velar) and voicing (voiced vs. voiceless), in order to guarantee ample phonemic variation. Moreover, the CVs consisted of the orthogonal combination of six consonants (/b/, /d/, /g/, /k/, /p/, /t/) and two vowels (/a/, /i/). A male talker while intending to convey a neutral affect produced all CVs. Amplitude contours (from Experiment 1) were imposed on the neutral CV productions using Syntrillium's Cool Edit Pro (1999). Amplitude was ramped linearly over the final 20 ms of each CV and normalized to peak amplitude.

A corresponding set of stimuli also was created that equated tokens for average root mean square (RMS) amplitude. Average RMS amplitude refers to the overall

loudness of a stimulus. Each token's overall loudness was adjusted so that all the stimuli were the same overall amplitude. The purpose of this manipulation was to correct for overall loudness differences that existed between the joy and sad contours. That is, stimuli imposed with the joy contour were louder overall than stimuli imposed with the sad contour. Equating the stimuli for average RMS thus reduced overall loudness differences, while still preserving the relative shape of the joy and sad contours. Thus, the equated version more directly examined whether overall amplitude was driving discrimination performance and/or if listeners were capable of utilizing contour information, as well.

Procedure

Listeners performed a two-alternative forced choice (2AFC) task, in which they were asked to identify which of two syllables on a trial reflected an amplitude contour consistent with either joy or sadness. On each trial, listeners were presented with identical syllables, one with a joy contour and one with a sad contour imposed. Each syllable and contour was equally likely to occur in first position, with the trials block randomized to six repetitions.

Each participant performed two versions of the 2AFC procedure: unequated and equated for RMS amplitude. Order of version of the task also was counterbalanced. There were two blocks of 48 familiarization trials. In a block of familiarization, two repetitions of each stimulus were presented across 24 trials. There were two blocks of 144 randomized experimental trials. Thus, in an experimental block there were 72 trials, which consisted of six trials for each syllable that could occur. All remaining aspects of

the stimulus presentation (e.g., filtering, intensity, method of delivery) were identical to those used in Experiment 2.

Design and Analysis

The dependent variable measured was d'. D' is a measure of sensitivity from Signal Detection Theory that is theoretically free of response bias (Green & Swets, 1966; MacMillan & Creelman, 1990). D' data was submitted to a 2 (Task Version) x 2 (Voicing) x 2 (Vowel) x 3 (Place of Articulation) repeated measures ANOVA, to compare the level of discrimination between joy and sad amplitude contours. The two levels of task version were unequated and equated, the two levels of voicing were voiced (/b/, /d/, /g/) and voiceless (/p/, /t/, /k/), the two levels of vowel were /a/ and /i/, and the three levels of place of articulation were bilabial, alveolar, and velar.

Results

In order to compare the overall level of discrimination between joyful and sad amplitude contours, d' scores were computed for each listener. D' scores were adjusted downward by a factor of the square root of two. This is a commonly used correction to account for the discrepancy in task difficulty/performance between 2AFC and yes-no tasks (MacMillan & Creelman, 1990).

In the *d'* analysis, a hit was defined as a response of joy-first when joy appeared in the first interval on a trial. A false alarm was defined as a response of sad-first on a trial where joy appeared in the first interval. Table 3 depicts individual and mean *d'* scores for listeners, in both the equated and unequated versions of the 2AFC task. As can

be seen in Table 3, the mean *d'* scores in the unequated version of the task differed significantly from the mean *d'* scores in the equated version. This difference was confirmed by a main effect for Task Version, $F(1, 9) = 40.329$, $p < 0.001$. Listeners were reliably able to discriminate joy contours from sad contours across trials in the unequated version of the task. This finding suggests that listeners were consistently able to use differences in overall amplitude as a cue, to discern between joy and sad contours.

In contrast, the majority of listeners in the equated version did not discriminate contour differences (see Table 3). Thus, without the presence of overall loudness differences as cues, listeners tended not to be able to rely solely on dynamic contour changes to identify joyful contours from sad contours.

Table 3 Individual *d'* scores (with means and standard errors in bold) for both the equated and unequated versions of the 2-AFC task

Unequated Version

Listener	ba	da	ga	ka	pa	ta	bi	di	gi	ki	pi	ti
1	3.28	4.66	3.28	3.28	4.66	4.66	3.28	4.66	4.66	4.66	4.66	4.66
2	4.66	3.28	4.66	2.77	3.28	3.28	3.28	3.28	2.77	3.28	4.66	4.66
3	4.66	4.66	3.28	3.28	3.28	1.39	0.95	3.28	0.88	2.77	1.91	1.39
4	4.66	1.91	0.95	4.66	0.00	2.33	2.77	1.91	1.39	2.33	4.66	3.28
5	4.66	2.77	1.91	2.77	3.28	4.66	3.28	1.91	3.28	3.28	1.39	2.77
6	4.66	4.66	4.66	4.66	4.66	2.77	4.66	3.28	4.66	3.28	3.28	3.28
7	4.66	4.66	4.66	4.66	2.77	4.66	4.66	4.66	3.28	4.66	4.66	4.66
8	3.28	1.91	4.66	3.28	3.28	1.89	2.77	4.66	3.28	3.28	3.28	4.66
9	3.28	4.66	4.66	4.66	3.28	4.66	4.66	4.66	4.66	4.66	3.28	4.66
10	4.66	4.66	4.66	4.66	4.66	4.66	4.66	4.66	4.66	4.66	4.66	4.66
M	**4.24**	**3.78**	**3.74**	**3.87**	**3.31**	**3.49**	**3.50**	**3.69**	**3.35**	**3.68**	**3.64**	**3.87**
SE	**0.21**	**0.38**	**0.43**	**0.27**	**0.43**	**0.42**	**0.38**	**0.36**	**0.44**	**0.28**	**0.39**	**0.36**

Equated Version

Listener	ba	da	ga	ka	pa	ta	bi	di	gi	ki	pi	ti
1	3.28	1.91	3.28	4.66	2.33	2.77	1.91	4.66	4.66	3.28	3.28	2.77
2	2.77	3.28	4.66	3.28	3.28	4.66	0.51	4.66	1.39	1.91	2.77	3.28
3	0.51	-1.39	-0.44	-0.88	-0.44	0.00	0.00	0.00	0.44	-0.44	-1.39	0.44
4	0.00	0.00	1.89	2.33	0.00	-0.44	-1.37	-0.44	-1.38	-1.38	0.00	-1.89
5	0.51	0.44	0.00	0.95	-0.95	1.39	3.28	1.91	1.91	4.66	0.44	2.33
6	0.95	-0.44	-0.44	1.39	-2.33	0.00	0.00	1.89	1.39	-0.95	0.00	0.44
7	0.00	0.00	1.37	0.00	-1.38	0.00	0.00	1.89	-0.44	0.00	-0.44	1.89
8	0.00	2.33	1.39	0.95	0.44	0.51	0.44	0.00	0.95	0.00	0.95	0.44
9	0.44	0.44	2.33	0.95	0.95	1.89	-0.88	3.28	0.51	0.51	0.44	0.95
10	-0.95	-0.95	0.00	0.95	0.00	1.39	-0.51	-0.88	-0.44	0.00	1.37	-2.77
M	**0.75**	**0.56**	**1.40**	**1.46**	**0.19**	**1.22**	**0.34**	**1.70**	**0.90**	**0.76**	**0.74**	**0.79**
SE	**0.41**	**0.47**	**0.54**	**0.50**	**0.53**	**0.50**	**0.43**	**0.64**	**0.52**	**0.61**	**0.45**	**0.61**

Sensitivity was influenced slightly by the place of articulation of the consonant stimuli. For example, ANOVA results revealed a nonsignificant main effect of Place, $F(2, 18) = 3.160, p > 0.07$, such that mean sensitivity was slightly higher for alveolar consonant stimuli ($d' = 2.396$) relative to velars ($d' = 2.388$) and lowest for bilabials ($d' = 2.090$). A Task x Place interaction, $F(2, 18) = 3.671, p < 0.05$ also was obtained (Mauchly's sphericity assumption was met). Specifically, the difference in performance across the unequated and equated versions of the task changed with place of articulation. The difference in sensitivity across tasks was greatest for the bilabials (mean difference in d' of 3.171), moderate for velars (mean difference in d' of 2.645), and least for alveolars (mean difference in d' of 2.531). More importantly, however, sensitivity was

significantly greater in the unequated version, regardless of place of articulation [as indicated by simple effects: bilabials, $F(1, 18) = 79.363, p < 0.001$; velars, $F(1,18) = 55.217, p < 0.001$; alveolars, $F(1, 18) = 50.560, p < 0.001$].

Interestingly, ANOVA results exposed a Place x Vowel x Voicing interaction, $F(2, 18) = 3.646, p < 0.05$. This finding indicates that the degree of change in mean sensitivity across place of articulation (bilabial, alveolar, velar) was influenced according to the combination of vowel and voicing. However, there were no systematic patterns of change in sensitivity levels between places of articulation, vowel, and voicing. For instance, mean sensitivity was greater for voiced bilabial consonants produced with the vowel /a/ ($M = 2.499$) relative to the voiceless version ($M = 1.753$). In contrast, sensitivity decreased for velars and alveolars at the same variable levels (velars, $M = 2.173$; alveolars, $M = 2.571$). There is no basis in the literature for predicting this three-way interaction. It is likely that the observed interaction is due to artifact and would not occur with a different set of subjects.

The magnitude of change in sensitivity varied across tasks according to the combination of vowel and voicing. This tendency was reflected in a three-way interaction for Task x Vowel x Voicing, $F(1, 9) = 7.214, p < 0.03$. The size of the difference in sensitivity between equated and unequated versions of the task decreased from voiced to voiceless consonants for /a/ vowels (mean difference in d' of 3.016 versus 2.605), but increased from voiced to voiceless consonants for /i/ vowels (mean difference in d' of 2.537 versus 2.970). In all conditions, however, there was still substantially greater sensitivity in the unequated version of the task. Thus, regardless of the combination of

vowel and voicing, listeners still relied on overall amplitude differences to classify stimuli as joy or sadness.

Closer examination of the data, however, shows that some listeners could actually discriminate dynamic patterns of amplitude in the absence of overall loudness. Three of the ten listeners demonstrated the capacity to detect differences between the contours in the RMS equated version of the task. Interestingly, these individuals had the most exposure to the contour patterns (i.e., they were involved in imposing the contours measured in Experiment 1 onto the PB words used in Experiment 2). This unexpected finding raises the possibility that if given more time and/or experience with the contours, listeners may be capable of representing dynamic changes in amplitude. This parallels the findings from talker training studies (e.g., Nygaard & Pisioni, 1992; Nygaard et al., 1994), which have demonstrated that increased experience with a talker improves word identification performance.

In summary, the primary goal of Experiment 3 was to determine whether listeners could reliably perceive the amplitude contours used in Experiment 2. The present results confirm that the amplitude manipulation in Experiment 2 was perceptually salient. Listeners were able to reliably distinguish joyful from sad amplitude contours on the unequated version of the 2AFC task, which consisted of an amplitude manipulation that was equivalent to that used in Experiment 2. This finding should not be considered surprising, insofar as the sadness contour was adjusted over an 18 dB range, which is equal to changing amplitude by a factor of three within a given utterance. After all, classic studies in psychophysics have shown that small changes in amplitude (on the order of 2 dB) can be accurately discriminated by listeners (Fletcher & Munson, 1933).

The fact that discrimination performance on the unequated version of the task was far superior to that obtained for the equated version of the task further indicates that listeners predominantly relied upon differences in overall amplitude to distinguish joyful from sad amplitude contours. However, the performance of three listeners on the equated version of the task suggests that with additional exposure, listeners may become more proficient at utilizing dynamic changes in amplitude, as a cue for signaling emotional contrasts. When taken collectively, these findings have important implications for the nature of representations for amplitude. These implications are discussed below.

CHAPTER 7

GENERAL DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

General Discussion

The goal of this study was to investigate the extent to which detailed, instance-specific information for spoken words are encoded and represented. The results that emerged from this study complement and extend the findings of earlier studies that have investigated the effects of talker and amplitude variability on speech perception and memory for spoken words.

The results from Experiment 1 demonstrate that people produce specific amplitude contours for words when they intend to convey the emotions of joy and sadness. This emotion-specific difference in amplitude contour is parsimonious with previous findings for static (i.e., overall) amplitude cues on the vocal expression and perception of emotion. For example, Scherer's (1986) observed that speech produced in a joyous manner tended to be produced with greater mean amplitude relative to vocal expressions of sadness.

Given such systematic differences in amplitude contours across emotion, it is plausible that this dynamic information serves a useful purpose. Amplitude contour information may convey additional meaning beyond that which is contained in the basic linguistic content of the utterance. Specifically, listeners may be sensitive to and

therefore, capable of using amplitude contour information as a potential cue in extracting emotion. In fact, such perceptual cues have already been established for static (i.e., overall) differences in amplitude across words. For example, Nygaard and Lunders (2002) demonstrated that listeners are sensitive to differences between RMS amplitudes for joyful and sad word productions.

The remaining experiments of the current investigation likewise sought to determine whether or not corresponding effects could be obtained for the time-varying differences in amplitude that were observed in Experiment 1. Experiment 2 specifically sought to address whether amplitude contour information associated with an emotion impacts recognition memory for a given word, and thus could be argued to be part of that word's representation. The data from Experiment 2 instead indicate that word recognition performance was unaffected by the dynamic manipulation of amplitude. This suggests that amplitude contours consistent with emotional expressions are not incorporated into lexical representations along with linguistic content. In contrast, talker variability reliably reduced word recognition performance (regardless of whether or not amplitude contour remained the same across repetitions of a given word). Thus, as with previous studies (e.g., Bradlow et al., 1999, Experiment 1; Church & Schacter, 1994; Goldinger et al., 1991; Nygaard et al., 1995; Palmeri et al., 1993; Schacter & Church, 1992; Sommers et al., 1994), Experiment 2 produced evidence that talker information is bundled with the representations of words.

The lack of amplitude effects on recognition memory in Experiment 2 is consistent with Bradlow et al.'s (1999) findings, in which their manipulation of overall amplitude failed to affect word recognition performance. Bradlow and her colleagues

have suggested that the extent to which different sources of variability are represented seems to depend on the specific source of stimulus variability, as well as the task and encoding conditions. Moreover, these researchers maintain that the varying degrees to which talker and amplitude variability affect speech perception and memory for spoken words may be due to differences in the relevance of each source of variability for the perception of phonetic contrasts. Variability in talker characteristics has been shown to have a significant impact on speech perception. For example, Ladefoged and Broadbent (1957) found that vowel identification could be altered depending on the perceived talker characteristics of a precursor phrase, and Johnson (1990) showed that perceived talker identity plays an important role in F0 normalization for vowels. Similarly, several studies have demonstrated rate dependencies for the processing of both vowels and consonants (e.g., Miller & Volaitis, 1989; Summerfield, 1981). In contrast, there are relatively few demonstrations of amplitude effects on phoneme perception, primarily being limited to examples of time-intensity trading in the perception of voicing contrasts (e.g., see Repp, 1982). Furthermore, experimental work on the suprasegmental (prosodic) characteristics of spoken language suggests that amplitude cues to phrase boundaries are not as salient as other types of cues. For example, Streeter (1978) found that listeners reliably used both the pitch contour and the duration pattern in parsing ambiguous algebraic expressions. By comparison, amplitude was observed to be used only in combination with appropriate values of duration.

The results in Experiment 2 are not consistent with a strong version of the exemplar-based approach to lexical representation (see Goldinger, 1996; Goldinger et al., 1991; Pisoni, 1997). This view would predict that all aspects of surface form are encoded

and represented. Differential effects could be found if lexical traces were assumed to include only some selected aspects of surface form. That is, representations of individual instances might not represent every perceptual dimension equally. Instead, some surface details might be included in lexical instance-based representations based on relevance of individual dimensions to the linguistic event.

According to this account then, amplitude information would only be stored in instances where it is used as a cue to phoneme perception. It is possible that the amplitude manipulation employed in the present investigation, which was emotional in nature, may be more important for higher order (cognitive) processes, such as determining the intended overall meaning of an utterance (e.g., discerning between sarcasm and joking), rather than phoneme perception. If this is the case, then one would not expect that amplitude contours which convey emotion are represented along with lexical information because the contours do not facilitate the perception of phonemes.

The failure to obtain evidence that dynamic changes in amplitude associated with emotion are represented with word information should not be taken as an automatic endorsement of the notion that all time-varying characteristics of amplitude are represented independently from word information. For instance, aspiration amplitude prior to voicing is known to contribute to the perception of voiced and voiceless stop consonants. Given that this aspect of amplitude variation therefore must necessarily affect word identification, it could be reasonably argued that aspiration amplitude could still be bound with the representation of a word. According to this perspective, any cue that is used to identify phonemes within the word (e.g., F0, rate, aspiration amplitude) should be represented along with word information. All other information, including prosodic cues

to phrase structure, amplitude contour information associated with emotion, and overall amplitude would either be discarded (via memory decay) or represented independently of word information. This would constitute a hybrid of exemplar theory in that not all aspects of the signal are represented for a word; only those parameters that contribute to word identification are stored. This perspective is essentially consistent with Bradlow et al. (1999) except that it allows for phonetically relevant aspects of amplitude to be represented with the word. Although amplitude contour associated with emotion does not appear to be represented with word information, it still appears to reflect perceptually salient information that can be represented independently of word information. This was demonstrated by the results of Experiment 3. In the unequated version of the 2AFC task, in which stimuli were not adjusted for overall amplitude (as with the stimuli in Experiment 2), all listeners *d'* scores were well above one. These scores indicate accurate discrimination of differences between joyful and sad contours.

In contrast, the majority of listeners in the equated version of the task, where overall amplitude differences did not exist displayed poor discrimination performance. Thus, listeners in the equated version generally could not use strictly amplitude contour information as a cue to discern between joyful and sad productions. Interestingly, however, three listeners in the equated version of the 2AFC task demonstrated reliable discrimination performance. The listeners that could accurately discriminate contour differences had substantial exposure to the contours. These individuals were involved in creating stimuli for Experiment 2, which required them to impose joyful and sad contours over several words. The benefit of their experience was not merely the result of increasing familiarity with a particular talker, but rather the consequence of exposure to

the specific amplitude contours. One possible explanation for the increased discrimination performance in these listeners' is that stimulus generation provided listeners' the opportunity to receive feedback with regard to the contour type. Given that individuals were required to listen to each stimulus following its creation, and that they were completely aware of the contour they had just imposed, these listeners essentially received immediate feedback with regard to the contour. It is likely that the feedback directed listeners' attention to specific contour differences, which facilitated the appropriate encoding and storage of information specific to each contour. Thus, this finding suggests that with sufficient exposure listeners may become capable of extracting amplitude contour information from the signal to form distinct, long-term representations of emotionally relevant cues.

This argument is consistent with other existing arguments that (at least some aspects of) voice and linguistic information are represented independently. For example, Schacter and Church (1992) proposed a theoretical framework that posits the existence of a presemantic auditory perceptual representational system (PRS). The PRS is presumed to handle information about auditory word forms separately from semantic information. The PRS is composed of a number of subsystems that process information about the form and structure, but not the meaning and associative properties of words, objects, and other types of stimuli (Church & Schacter, 1994; Schacter & Church, 1992).

Similarly, Ferreira (1993) and others researchers (see also Bock & Loebell, 1990; Ferreira, 1991; Gee & Grosjean, 1983; Levelt, 1989) have argued for a distinctly prosodic level of representation, in which the prosodic structure or the sound pattern of an utterance is represented independent of linguistic information. According to Ferreira

(1993), prosodic structure is created from a sentence's syntactic structure but without knowledge of the phonemic content.

Neuropsychological research also suggests the existence of a form versus semantic dissociation within the auditory domain. Specifically, a number of studies have described patients who are able to repeat spoken words but do not understand them, a condition referred to as word meaning deafness. In cases of word meaning deafness (Ellis, 1982), patients can exhibit access to the meaning of words through other modalities, as indicated by their normal reading comprehension and use of words in spontaneous speech. There is also some evidence that such patients can write words to dictation as well as repeat them, thereby suggesting that they can gain access to stored auditory word form representations (Ellis, 1982). Thus, these patient's deficits are produced by a disconnection between a normally functioning system that handles acoustic and phonological properties of spoken words and a normally functioning semantic system. In cases of transcortical sensory aphasia, which is caused by damage to fibers surrounding Wernicke's area, patients can recognize words and they can talk, but they cannot understand what people are saying to them and have no spontaneous speech of their own (Coslet, Roeltgen, Rothi, & Heilman, 1987). This pattern of behavior suggests damage to semantic (i.e., language-based) systems despite an intact general auditory system. Similarly, the major findings from the current investigation support the notion of separate representations of semantic information for words and general auditory information from amplitude contours associated with emotion.

General Conclusions and Future Directions

An important motivation for the current investigation was to understand both the sources of variability in the speech signal and the effects of this variability on the listeners' word identification and recognition. Experiment 1 established that amplitude contours are a potentially meaningful source of variability contained within the speech signal, as indicated by the fact that talkers produced distinct amplitude contours for words spoken with the intent to convey the emotions joy and sadness. This production research extends our current knowledge of the vocal expression of emotion by providing quantitative data as opposed to the previously available descriptive accounts of how amplitude changes as a function of affect (Scherer, 1986). Prior research also has predominantly focused on how talkers modify only certain vocal parameters, such as fundamental frequency, speaking rate, and overall amplitude in expressing emotion (Bradlow et al., 1999; Scherer, 1986). The current investigation thus adds to this emotion literature by providing data on systematic, dynamic changes in talker amplitude.

While it is clear that talkers produce distinct amplitude contours for joy and sadness, questions remain regarding the representation of dynamic fluctuations in amplitude. Given that amplitude contours did not affect word recognition performance in Experiment 2, it appears that dynamic changes in amplitude that provide relevant emotional information may not be bound with the representation of other (phonetic) information about a word. The fact that some listeners in Experiment 3 were proficient at distinguishing between joy and sad amplitude contours in the absence of loudness differences across utterances suggests that listeners can represent such contour information.

This is a new observation in the literature, in that it shows that direct experience and/or training with a specific form of variability, not just mere exposure to a particular talker's voice, may improve the efficiency with which contours are identified. This suggests that the speech perception mechanism is susceptible to processes of learning and attention and that amplitude contour information is capable of being stored in some fashion.

The effects of experience on the representation of amplitude contours suggested here should be easily elucidated by a training study. Toward this end, ongoing research in our laboratory expects to systematically examine the effect of amount of training and feedback on listeners' ability to identify differences between joy and sad contours. Findings from the training study should confirm whether or not amplitude contour information is represented. If guided practice influences a listener's ability to discriminate joy from sad contours then this would suggest that experience provides an opportunity to construct more stable contour representations. Subsequently, the impact of training and feedback on word recognition performance also will be examined to validate the results obtained in Experiment 2, which potentially suggest that amplitude contour information is distinctly represented from phonetic information. If the training procedure is found to affect listeners' word recognition performance, then this would reveal that the mechanisms responsible for the encoding of talker information are linked directly to those that underlie phonetic perception.

Another means of addressing amplitude's relative importance in word recognition, and thus, representation, is to examine the time-varying relationship between aspiration amplitude and voicing. As previously demonstrated by Repp (1982), there

exists a time-intensity trade-off between amplitude and voicing, which influences listeners' phoneme perception. Using the PB word lists from Experiment 2, we will manipulate words that contain either an initial or final voicing contrast (i.e., any word that begins or ends with the consonant /b/, /d/, /g/, /k/, /p/, or /t/). The manipulation will involve an increase of aspiration amplitude for words that begin or end with a voiceless consonant, and conversely, a decrease of amplitude by a corresponding amount for words that begin or end with a voiced consonant. Listeners will perform a word recognition task identical to the one used in Experiment 2. If word recognition performance is negatively impacted by changes in aspiration amplitude across repetitions of a word, then this would suggest that dynamic changes in amplitude may be bound with the representation of other phonetic information about a word. However, if memory performance is unaffected, then this would provide corroborating evidence (with Bradlow et al., 1999, along with findings from the current investigation) that amplitude is a linguistically irrelevant vocal parameter, and accordingly, is not represented.

In summary, future research seeks to resolve whether dynamic changes in amplitude play a role in the recognition and representation of spoken words. The current investigation clearly demonstrates that talkers produce systematic changes in amplitude contour when expressing emotion. Such contour information does not appear to be incorporated in the representations of words insofar as it was not found to affect word recognition performance. Subsequent tests of discrimination performance indicated that listeners primarily rely on overall loudness differences rather than dynamic fluctuations in amplitude to discern between joyful and sad amplitude contours. Additionally, there were some indications that these amplitude contours may be independently represented .

Specifically, listeners who inadvertently received exposure to the contours demonstrated increased sensitivity in distinguishing between joyful and sad amplitude contours in the absence of overall loudness differences. Future research will address the contribution of experience to the formation and nature of representations for amplitude contour, as well as examine whether or not other dynamic aspects of amplitude, such as aspiration amplitude are linguistically relevant, and therefore, potentially incorporated in word representations. When taken together with the major findings from the current investigation, a more thorough understanding of 1) the exemplar-specific nature of word representations, and 2) the representation of amplitude-based cues to auditory affect should be achieved.

REFERENCES

American National Standards Institute, (1971). Method for measurement of monosyllabic word intelligibility (American National Standards S3-2-1960 [R1971]). New York: Author.

Arnold, M. B. (1960). Emotion and personality: Volume I Psychological Aspects. New York: Columbia University Press.

Bock, J. K. & Lobell, H. (1990). Framing sentences. *Cognition, 35*, 1-40.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics, 61*(2), 206-219.

Carrell, T. D. (1984). Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification. *Research on Speech Perception* (Technical Report No. 5). Bloomington: Indiana University, Department of Psychology.

Church, B. A. & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(3), 521-533.

Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research, 14*, 565-577.

Coslet, H. B., Roeltgen, D. P., Rothi, L. G., & Heilman, K. M. (1987). Transcortical sensory aphasia: Evidence for subtypes. *Brain and Language, 32,* 362-378.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89*, 627-661.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). Emotion in the human face: Guidelines for research and an integration of findings. New York: Pergamon Press (2nd ed., P. Ekman Ed.), Cambridge, England: Cambridge University Press.

Ellis, A. (1982). Modality-specific repetition priming of auditory words recognition. *Current Psychological Research, 2*, 123-128.

Fant, G. (1960). Acoustic theory of speech production. Mouton: The Hague.

Ferriera, F. (1993). Creation of prosody during sentence production. *Psychological Review, 100*, 233-253.

Ferriera, F. (1991). Effects of length and syntactic complexity on initial times for prepared utterances. *Journal of Memory and Language, 30*, 210-233.

Fletcher, H. & Munson, W. (1933). Loudness, its definition, measurement, and calculation. *Journal of the Acoustical Society of America, 5*, 82-108.

Garner, W. R. (1974). The processing of information and structure. Potomac, MD: Erlbaum.

Gee, J. P. & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*, 411-458.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 1166-1183.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*(1), 152-162.

Green, D. M. & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley.

Greenwood, J. D. (Eds.). (1991). The future of folk psychology: Intentionality and cognitive science. Cambridge: Cambridge University Press.

Haytar, A. J. (1986). The maximum familywise error rate of Fisher's least significant difference test. *Journal of the American Statistical Association, 81*, 1000-1004.

Hintzman, D. L. (1986). Schema abstraction in a multiple trace memory model. *Psychological Review, 93*, 411-421.

Izard, C. E. (1971). The face of emotion. New York: Appleton-Century Crofts.

Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America, 88,* 642-654.

Joos, M. A. (1948). Acoustic phonetics. *Language, 24*, 1-136.

Kendall, R. A. (2000). Music Experiment Development System, (Version 2001-A) [Computer Software]. Los Angeles, CA: Author, www.ethnomusic,ucla.edu/Faculty/Kendall/meds.htm.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*(2), 93-107.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language, 56,* 485-502.

Ladefoged, P. & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98-104.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.

Lieberman, P. (1961). Some acoustic measures of fundamental periodicity of normal and pathologic larynges. *Journal of the Acoustical Society of America, 35*, 344-53.

Lisker, L. & Abramson, A. S. (1967). Some effects of context on voice onset time in english stops. *Language and Speech, 10*, 1-28.

Lively, S. E. & Pisoni, D. B. (1997). On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 23*(6), 1665 1679.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America, 103*(6), 3648-3654.

MacMillan, N. A. & Creelman, C. D. (1991). Detection theory: A user's guide. Cambridge University Press: Cambridge.

Martin, C. S., Mullenix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 676-684.

Medin, D. L. & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Miller, J. L. & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception and Psychophysics, 46*, 505-512.

Minda, J. H. & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 775-799.

Mullenix, J. W. & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*(7), 379-390.

Mullenix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Mullenix, J. W. & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*(4), 379-390.

Nosofsky R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Nosofsky, R. M. & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 908-923.

Nosofsky, R. M. & Zaki, S. R. (1998). Dissociations between categorization and recognition memory in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 247-255.

Nygaard, L. C. & Lunders, E. R. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition, 30*, 583-593.

Nygaard, L. C. & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*, 355-376.

Nygaard, L. C., Sommers, M.S., & Pisoni, D. B. (1994). Speech perception as a talker contingent process. *Psychological Science, 5*(1), 42-46.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57*(7), 989-1001.

Ortony, A. & Turner, T. (1990). What basic about basic emotions? *Psychological Review, 97*, 315-331.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(2), 309-328.

Palmeri, T. J. & Nosofsky, R. M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54A*, 197-235.

Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In J. Mullenix & K.A. Johnson (Eds.), *Talker variability in speech processing* (pp. 9 32). New York: Academic Press.

Pisoni, D. B., Miyamoto, C., Kirk, K. I., Sommers, M., & Osberger, M. J. (1994). Sources of variability in speech perception hearing impaired listeners. *Progress Report No. 19, National Institutes of Health.*

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion* (pp. 3-33). New York: Academic.

Rand, T. C. (1971). Vocal tract size normalization in the perception of stop consonants. *Haskins Laboratory Status Report Speech Research. SR-25/26, 141-146.*

Reilly, S. S. & Muzekari, L. H. (1979). Responses of normal and disturbed adults and children to mixed messages. *Journal of Abnormal Psychology, 88*, 203-208.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*, 81-110.

Rosch, E. H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 3*, 192-233.

Rosch, E. H. (1978). Principles of categorization. In: E. Rosch & B. Lloyd, (Eds.), *Cognition and Categorization.* Hillsdale, N.J.: Erlbaum Associates.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110-114.

Schacter, D. L. & Church, B. A. (1992). Auditory priming and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 915-930.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143-165.

Smith, J. D. & Minda, J. P. (2001). Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 984-1002.

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude. *Journal of Acoustical Society of America, 96,* 1314-1324.

Streeter, L. A. (1978). Acoustic determinants of phase boundary perception. *Journal of Acoustical Society of America, 64,* 1582-1593.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1074-1095.

Summerfield, Q. (1975). Towards a detailed model for the perception of voicing contrasts. *Speech Perception Report No. 3* (Queens University, Belfast).

Sussman, H. M. (1986). A neuronal model of vowel normalization and representation. *Brain & Language, 28,* 12-23.

Syntrillium Software Corporation, Cool Edit Pro [Computer software] (1999).

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 2,* 339-363.

Verbrugge, R. R., Strange, W., Shakweiler, D. P., & Edman, T. R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America, 60,* 198-212.

Viemeister, N. F. & Bacon, S. P. (1982). Forward masking by enhanced components in harmonic complexes. *Journal of the Acoustical Society of America, 71,* 1502-1507.

Volaitis, L. E. & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America, 92*, 723-735.

Zaki, S. R. & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1022-1041.

VITA

Graduate College
University of Nevada, Las Vegas

Kimberly M. Cramer

Home Address:
301 Butterworth Court
Henderson, NV 89052

Degrees:
Bachelor of Science, Kinesiology, 1997
University of Nevada, Las Vegas

Master of Science, Kinesiology, 2001
University of Nevada, Las Vegas

Publications:

Hall, M. D., & Wieberg, K. M. (2003). Illusory conjunctions of musical pitch and timbre. *Acoustic Research Letters Online, 4*(3), 65-70.

Hall, M. D., & Wieberg, K. M. (2002). Distinguishing feature misperception from illusory conjunctions in spatially distributed musical tones. *Journal of the Acoustical Society of America, 112*(5), 2274-2287.

Hall, M. D., & Wieberg, K. M. (2000). The role of timbre similarity in the binding of musical features. *Journal of the Acoustical Society of America, 108*(5), 2642 2653.

Thesis Title: The Impact of Dynamic Changes in Talker Amplitude on Recognition Memory for Words

Thesis Examination Committee:
Chairperson, Dr. Michael Hall, Ph.D.
Committee Member, Dr. Gretchen Kambe, Ph.D.
Committee Member, Dr. Ned Silver, Ph.D.
Graduate Faculty Representative, Dr. David Beisecker, Ph.D.