PROSODIC PERSISTENCE IN MUSIC PERFORMANCE
AND SPEECH PRODUCTION

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Melissa Kay Jungers, M.A.

The Ohio State University
2003

Dissertation Committee:

Dr. Caroline Palmer, Adviser

Dr. Neal Johnson

Dr. Mark Pitt

Dr. Shari Speer

Approved by

Adviser

Psychology Graduate Program

ABSTRACT

Prosodic cues play a similar role in speech and music; they can differentiate interpretations of ambiguous sentences or melodies and they can aid memory. Prosodic cues can be related to the syntactic structure of an intended production, but they can also exist separately from the structural features. Four experiments examined the persistence of structurally-related and structurally-unrelated prosodic cues in production. In Experiment 1, pianists listened to melodies containing intensity and articulation cues and performed similar metrically ambiguous melodies. Pianists' performances persisted only in the metrically-unrelated articulation cues, but in a metrically related manner. In Experiment 2, speakers heard sentences containing prosodic phrase breaks and tonal patterns and produced similar syntactically ambiguous sentences. Their productions reflected syntactically-related prosodic phrase breaks and some evidence of syntactically-unrelated tonal (pitch) accents. Experiments 3 and 4 examined listeners' ability to judge the meter or syntax from the productions collected in Experiments 1 and 2. Listeners correctly identified the meter and syntax of the productions. These results suggest that prosodic persistence may be beneficial in a conversational or ensemble context.

Dedicated to my wonderful family

ACKNOWLEDGMENTS

I wish to thank Caroline Palmer for her encouragement and support over the past five years. Merci!

I also wish to thank Shari Speer for her enthusiastic introduction to the world of psycholinguistics.

I thank Neal Johnson and Mark Pitt, who offered their insight and suggestions.

I wish to thank Laurie Maynell for her help and advice in the creation of stimuli.

I am also grateful to Beth Mechlin, Lindsay Barber, Chris Hanson, Jenna Johnson, and Michael Keida, who ran subjects and entered data.

I thank the Cognitive Science Center at Ohio State University for supporting my research and connecting me with two great collaborators.

VITA

1998.....B.S. Psychology, Bowling Green State University
2000.....M.A. Psychology, Ohio State University
1998-present.....Graduate Fellow and Research Assistant
The Ohio State University

PUBLICATIONS

Research Publication

1. Jungers, M.K., Palmer, C., & Speer, S.R. (2002). Time after time: the coordinating influence of tempo in music and speech. *Cognitive Processing*, 2, 21-35.
2. Palmer, C. & Jungers, M.K. (2001). Music cognition. R. Goldstone (Ed.), *Encyclopedia of Cognitive Science*. London: Macmillan.
3. Palmer, C., Jungers, M.K., & Jusczyk, P. (2001). Episodic memory for musical prosody. *Journal of Memory and Language*, 45, 526-545.

FIELDS OF STUDY

Major Field: Psychology

TABLE OF CONTENTS

	<u>Page</u>
Abstract.....	ii
Dedication.....	iii
Acknowledgments.....	iv
Vita.....	v
List of Tables.....	vii
List of Figures.....	viii
Chapters:	
1 Introduction.....	1
2 Control Experiment 1: Music stimuli.....	19
3 Control Experiment 2: Speech stimuli.....	26
4 Experiment 1: Music production.....	33
5 Experiment 2: Speech production.....	43
6 Experiment 3: Music perception.....	58
7 Experiment 4: Speech perception.....	64
8. General Discussion.....	68
List of References.....	77

LIST OF TABLES

<u>Table</u>	<u>Page</u>
3.1 Control Experiment 2 & Experiment 2: Four types of syntactically ambiguous sentences with interpretations.....	29
5.1 Experiment 2: ToBI transcription for prime utterances.....	49
5.2 Experiment 2: Percentage of target utterances by participants following Early or Late prime sentences that contain pitch accents or phrase breaks at specified locations.....	51
5.3 Experiment 2: Chi-square table for prime and target utterances.....	53

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1 Control Experiment 1 & Experiment1: Sample metrically ambiguous melody.....	21
4.1 Experiment 1: Mean ISI/IOI for pianists' target performances following legato and staccato prime articulation	38
4.2 Experiment 1: Mean ISI/IOI for each event in pianists' target performances following binary and ternary prime melodies.....	40
4.3 Experiment 1: Mean percent 'yes' response for melody recognition task by cue condition.....	41
5.1 Experiment 2: Mean percent 'yes' response for sentence recognition task by cue condition.....	55
5.2 Experiment 2: Mean percent 'yes' response for same or different prosodic phrase breaks.....	56
6.1 Experiment 3: Mean percent 'ternary' response in musical perception task with intensity or articulation cue.....	61

CHAPTER 1

INTRODUCTION

When two people have a conversation, does the way in which one person produces sentences influence the way in which the second person speaks? When two musicians perform together, does the style of one musician influence the style of the other? Although these questions come from different domains, the underlying issue is the same. What prosodic dimensions persist among producers and listeners in speech and music?

The current study examines the persistence of prosodic cues in both speech and music. In order to persist in the prosody of the utterance or performance just produced, the listener must remember the prosody. How well do listeners retain acoustic features in memory? If listeners do not retain the acoustic details from a production, then they would not be expected to produce these details in subsequent productions.

Why might speakers and performers persist in the acoustic cues of what they have just heard? One possibility is that persistence aids communication in a conversational speech context or a small-ensemble musical context. Perhaps the two (or more) parties in a conversation mutually influence each other's prosody. This may make communicating a message easier because the parties focus on the message and are not distracted by the acoustic variation.

Prosody

What *is* prosody? Prosody has been described as a structure that organizes sound as well as the suprasegmental features of speech such as pitch, timing, and loudness (Cutler, Dahan, van Donselaar, 1997). Thus, prosody refers to both an abstract hierarchical structure and the fine acoustic details in speech, such as the “variation in fundamental frequency, spectral information, amplitude, and the relative duration of sound and silence” (Speer, Crowder, & Thomas, 1993). Prosody is also described as the “stress, rhythm, and intonation in spoken sentences” (Kjelgaard & Speer, 1999). Informally, prosody refers to the way in which something is spoken, rather than just the words.

Prosodic dimensions such as word duration, timing, and intonation influence listeners’ interpretation of sentence meaning. The specific acoustic details associated with a speaker’s voice can aid sentence interpretation (Nygaard & Pisoni, 1998). In one study, listeners were familiarized with ten speakers’ productions of isolated words (Nygaard & Pisoni, 1998). At test, listeners showed better intelligibility for the familiar voices than unfamiliar voices for isolated novel words in noise. Also, sentences produced by a familiar speaker were more intelligible to listeners than sentences produced by a new speaker (Nygaard & Pisoni, 1998). Word durations can disambiguate the meaning of ambiguous sentences (Lehiste, 1973; Lehiste, Olive, & Streeter, 1976). Listeners used acoustic features to determine the intended meaning of different versions of syntactically ambiguous sentences. Analysis of the acoustic properties from the sentences suggested that timing and intonation were useful features for disambiguation

(Lehiste, 1973). The placement and duration of pauses provides another perceptual cue to sentence meaning; speakers' pause patterns tend to correlate with the syntactic structure of a sentence, with longer pauses near important structural boundaries (Lehiste, et al., 1976). Prosodic emphasis also influences interpretation of sentence meaning (Speer, et al., 1993). Listeners heard syntactically ambiguous sentences that contained different prosodic realizations of a single word, such as the sentences "They are FRYING chickens" and "They are frying CHICKENS". Listeners' paraphrasings of the sentences showed that the interpretation depended on the prosodic emphasis. A diverse set of prosodic features, including duration, timing, and intonation, influence the interpretation of sentences.

In order to examine and discuss prosody in speech, a system is needed so that the speech signal can be consistently described. Beckman and Pierrehumbert (1986) designed this type of transcription system for describing the prosodic aspects of a spoken sentence. Their system is called ToBI, which stands for Tone and Break Indices. This system has four parallel tiers: orthographic, tone, break index, and miscellaneous (Pierrehumbert & Hirschberg, 1990). Certain tones are stressed and are known as pitch accents. English includes high (H) and low (L) tones. Every utterance contains an intonational phrase that is delimited on the right edge by a boundary tone (L% or H%). Each intonational phrase (Iph) is made up of at least one phonological (Pph) or intermediate phrase (ip) which is delimited on the right edge by a phrase accent (H- or L-). Each phonological phrase contains at least one pitch accent. Prosody is hierarchical, but this hierarchy is shallower than that of syntax (Cutler et al., 1997).

How do prosody and syntax relate? Syntax refers to the rules native speakers use to put words together into sentences. There is not a one-to-one mapping between prosody and syntax. However, prosodic boundaries often coincide with syntactic boundaries. For example, segmentation of a sentence relied on both syntactic structure as well as the acoustic pattern in a listening task (Wingfield & Klein, 1971). Subjects heard spliced sentences in which a complete phrase was put into a sentence that matched or did not match the intonation pattern of the phrase. The task was to determine the point at which the recorded sentence switched from one ear to the other. Both the syntactic form and the prosodic pronunciation influenced the determination for the switch. The authors concluded that segmentation is determined primarily by syntactic structure, but the acoustic pattern helps to mark the syntax (Wingfield & Klein, 1971).

Prosody also helps to disambiguate syntactically ambiguous sentences. In another early experiment, Lehisté (1973) asked four speakers to read grammatically ambiguous sentences. Listeners heard the sentences and had to guess the speakers' intended meaning. Listeners were better than chance for 10 of the 15 sentences (Lehisté, 1973). Analysis of acoustic properties from the sentences revealed that timing and intonation were successful strategies for disambiguation (Lehisté, 1973). Those sentences that were difficult for subjects to interpret were those sentences in which the difference in meaning did not correlate with a different surface constituent structure, for example, "John doesn't know how good meat tastes" (Lehisté, 1973).

Past research indicated that listeners use prosody to determine the meaning of an ambiguous sentence (Lehiste, 1973, Lehiste, et al., 1976), but several recent papers questioned whether this effect is partly a result of using trained speakers who produced intonation and timing patterns that clarify ambiguities for listeners. Do people produce and perceive prosodic cues to resolve syntactic ambiguity in normal conversations? In one study (Albritton, McKoon, & Ratcliff, 1996), trained and untrained speakers read syntactically ambiguous sentences that were embedded in two different passages that clarified the intended meaning. Two judges rated their intended meaning. When the untrained speakers were unaware of the ambiguity, they read the passages without disambiguating the embedded sentences, according to the two judges' ratings about the intended meaning. Likewise, trained speakers who were unaware of the ambiguity did not disambiguate the meaning. Only trained and informed speakers were judged, by both independent raters and by naïve listeners, to have disambiguated the meaning when reading the sentences (Albritton, et al., 1996). The authors concluded that although it is possible to use prosody to disambiguate syntax, prosody may be a relatively minimal cue and its use may not translate to conversational speech outside the laboratory (Albritton, et al., 1996).

In another study examining natural speech, speakers memorized and then produced short passages containing an embedded syntactically ambiguous sentence (Fox Tree & Meijer, 2000). The researchers pitted prosody against context by inserting either the original sentence or an incongruent middle sentence between two sentences that indicated a semantic context. The incongruent middle sentence contained the same

words as the original middle sentence, but came from a production of the sentence within another context. When listeners chose one of two intended meanings of the embedded sentence, they made their decision based on the context and not on the prosody of the embedded sentence (Fox Tree & Meijer, 2000). The authors concluded that prosodic cues are not consistent enough to use for syntactic disambiguation in everyday conversation. From this research, it may appear that prosodic cues are not generally useful for syntactic disambiguation in a conversational context. However, several problems in this study make this conclusion less clear. For example, the stimuli were created by naïve speakers who read the passages silently and then delivered them from memory. This method of production is more natural than reading, but speakers may have been more concerned with correctly saying the words from the three sentences than with communicating the idea of the passage. Also, the listeners' choice of intended sentence meaning could be made without even hearing the embedded sentence since the semantics of the context sentences made the answer clear. In addition, listeners were not told to use prosody (or even the middle sentence) to make their decision, so it is not surprising that they used the context.

Other research has demonstrated the use of prosody by untrained speakers in a natural context. In a game task, naïve speakers produced prosodic cues to disambiguate syntax (Schafer, et al., 2000). The set of possible utterances was limited, so the speakers learned to produce them without making errors or referring to the text. The speakers disambiguated the syntax with prosodic cues, even when the game situation did not require disambiguation (Schafer, et al, 2000).

The debate about the use of prosody to interpret ambiguous syntax in natural contexts continues, but there is certainly evidence to suggest a link between prosody and syntax (Wingfield & Klein, 1971; Lehiste, 1973, Schafer et al., 2000). The relationship between these speech elements is not isomorphic (Cutler et al., 1997). In fact, Beckman (1996) argues that prosody has its own structure that is parsed.

Musical prosody

What is the musical equivalent to prosody? Prosody is expressiveness in music; the acoustic features that performers add, beyond what composers specify in notation. Such features are referred to as performance “expression,” and can differentiate two performances of the same music (Palmer, 1997).

What is musical syntax and how does prosody relate to it? Western tonal music contains style-specific syntactic properties, such as meter and grouping (Cooper & Meyer, 1960, Lerdahl & Jackendoff, 1983). Meter refers to the alternation of strong and weak beats. For example, the beats in a march (2/4, 4/4) alternate between strong and weak while the beats of a waltz (3/4, 6/8), follow a strong, weak, weak pattern. Grouping is based on pitch relationships or rhythmic patterns (Lerdahl & Jackendoff, 1983; Cooper & Meyer, 1960). Both meter and grouping are hierarchically arranged, with sequences divided into smaller sequences of pitches or rhythms.

The acoustic features that contribute to performance expression are related in a rule-based way to the printed score. For example, differences between categories of length or pitch are exaggerated, such that short notes are shortened and long notes are lengthened (Sundberg, 1999). Also, small pauses are inserted between pitch leaps and

musical phrases, emphasizing the grouping structure of the music (Sundberg, 1999). Decreased tempo and dynamics are expected at the ends of phrases (Windsor & Clarke, 1997, Henderson, 1936). This phrase-final lengthening also indicates the hierarchical importance of the phrase (Lerdahl & Jackendoff, 1983; Palmer, 1996a; Palmer, 1997). Meter is also expressed through acoustic features. Events that align with metrically strong beats are performed with increased duration, louder accents, and a more legato articulation than weak beats (Sloboda, 1983, 1985). One study examined whether accents associated with different musical structures (meter, rhythmic grouping, and melodic accent) influence performance expression independently or interactively (Drake & Palmer, 1993). Meter and rhythmic grouping influenced performance expression independently, but the influence of melodic accent on expression depended on the context (Drake & Palmer, 1993).

The relationship between performance expression and structure also influences the way listeners perceive music. When listeners heard performances that were altered to contain one or more of the acoustic cues, listeners used the articulation cues to choose the intended meter (Sloboda, 1985). Loudness was also used to determine meter, but not all performers differentiated meter with loudness (Sloboda, 1985). Listeners were less likely to detect a computer-lengthened event before a long duration in a simple rhythmic pattern (Drake, 1993); the same location at which performers often lengthen events (Drake & Palmer, 1993). Also, listeners were less likely to detect a lengthened event in a computer-generated performance when it occurred at a structurally-expected location (Repp, 1992). Listeners' judgments of goodness of fit of a probe beat inserted in a

metrical context reflected knowledge of metrical accent structure (Palmer & Krumhansl, 1990). Thus, listeners use the prosodic features to determine the structure of a production and they also use the structure to guide their perception.

Memory for prosody

Human listeners can understand speech that is produced by men, women, and children, with different vocal pitch ranges. They can understand speakers with foreign accents and even speakers with colds. For this reason, many studies on speech perception focused on listeners' ability to ignore the prosodic details in order to understand the message. This approach of normalization assumes that the listener transforms the physical speech signal into a standard representation of the sentence devoid of prosodic details (Pisoni, 1997). It is this bland representation that is stored in memory. According to this approach, the acoustic details of a production will not be retained.

Several more recent studies suggest that acoustic features of speech are incorporated in memory for language. Sentences are recognized more accurately when they are presented with the same prosody at learning and test. Also, prosodic cues aid memory for syntactically ambiguous sentences (Speer, et al, 1993). Listeners can use extralinguistic information, including talker identity and talker's rate, to accurately identify previously presented words (Bradlow, Nygaard, & Pisoni, 1999). The rate of presentation affects listeners' abilities to recall items produced by different speakers. Listeners show better recall for those items presented at the same rate in both

familiarization and test than for items presented at different rates from familiarization to test (Nygaard, Sommers, & Pisoni, 1995). These findings suggest that prosodic cues influence memory for speech contents.

When you walk away from a concert humming, do you remember only the melody or do you remember the *way* in which the song was performed? Some early research in music perception focused on listeners' ability to recognize a tune, even when performed at a different tempo or with different instrumentation. This approach of perceptual constancy, in which the stimulus sounds the same although it is physically different, assumed that some of the acoustic details were removed in order to recognize the underlying similarity of a musical excerpt (Dowling & Harwood, 1986; Large, Palmer, & Pollack, 1995). Through a normalization process similar to the one proposed for speech, acoustic details were filtered out and only the underlying representation was stored in memory (Large, et al., 1995).

Another reason the prosodic cues may not be retained in memory is because listeners do not have fine-grained memories for music performances (Raffman, 1993). Raffman (1993) suggests prosodic cues are used to form categories of pitch and rhythm, but the specific acoustic details are not retained. There is evidence that even trained musicians do not accurately identify small within-category pitch differences (Siegel & Siegel, 1977). Thus, memory for music will be limited to larger pitch and rhythm categories.

Some studies demonstrate that listeners remember particular acoustic features from performances. Palmer, Jungers, and Jusczyk (2001) examined memory for acoustic details in music performance. Listeners with and without music training were familiarized with one of two performances of the same short musical excerpt. The performances differed in articulation, intensity, and interonset interval cues. At test, the listeners heard the original performances from familiarization as well as different performances of the same melodies (the same notated pitches and durations), but different intensities, articulations, and interonset intervals. Listeners were asked to identify which of the performances were present at familiarization. Listeners could recognize the performances of the melodies they had heard during familiarization, even though the categorical pitches and durations in the two versions were identical (Palmer et al., 2001).

The adult listeners in Palmer et al. (2001) had many years of exposure to music. To address whether this musical acculturation is necessary for memory for musical features, Palmer et al. (2001) tested 10-month-old infants' memory for performances with the same melodies, using a head-turn preference procedure (Kemler Nelson et al., 1995). After being familiarized with one performance of each melody, infants oriented longer to the familiar performances during test than to other performances of the same melodies. Thus, even infants (with little exposure to music) can use acoustic cues that differentiate performances to form a memory for short melodies (Palmer et al., 2001).

Although this study indicated that listeners are sensitive to subtle performance differences and can retain them in memory, it does not indicate which prosodic cues are most salient in perception and memory. In another study, musician listeners were tested for their ability to discriminate and remember music performances that differed in only one or two acoustic cues (Jungers & Palmer, 2000). In a discrimination task, listeners could accurately distinguish same from different pairs of performances of the same melody listeners when articulation or articulation with intensity cues were present. In a memory task, musician listeners were familiarized with performances that varied in articulation, intensity, or articulation with intensity cues and later heard these performances as well as novel performances of the same melody. Listeners could more accurately identify performances they had heard before and were most accurate at identifying those performances that varied in articulation cues (Jungers & Palmer, 2000). Thus, listeners were particularly sensitive to the articulation cues in music performances; listeners discriminated musical sequences based on the timing between pitch events within the sequence.

Both musician and non-musician listeners can remember particular performance tempi over prolonged timed periods. Musicians can reproduce performances of long musical pieces, such as an entire movement of a symphony, at the same tempo with very low variability (Clynes & Walker 1986; Collier & Collier, 1994). Similarly, nonmusicians can reproduce popular songs from memory at tempi very close to the original tempo (Levitin & Cook, 1996). Furthermore, when people sang familiar songs as fast or as slow as possible, songs that lacked a tempo standard in original recordings were

produced with a larger variability in tempo; this counters arguments that memory for the tempo of remembered songs was solely a function of articulatory constraints (Levitin & Cook, 1996).

In sum, listeners demonstrate the ability to remember prosodic details in both speech and music. Listeners recognize sentences more accurately when the same prosodic cues are present at learning and test (Speer et al., 1993). Listeners can use information such as talker identity and rate to identify previously heard words (Bradlow, et al., 1999). Adults and infants remember the specific acoustic details of a performance and can distinguish previously heard performances from different performances of the same categorical pitches and durations (Palmer, et al., 2001). Also, nonmusicians reproduce popular songs at the tempo they have heard before (Levitin & Cook, 1996).

Prosody details of language and music become part of the memory representation. Do these prosodic cues influence future performances? Several studies point to both acoustic and structural features that persist in speech and music.

Prosodic persistence

One aspect of speech that may persist is the rate. Kosslyn and Matt (1977) played a recording of two male speakers for listeners: one speaking at a fast rate and one at a slow rate. Then the subjects read a passage they were told was written by one of the speakers. The subjects imitated the rate of the speaker who supposedly wrote the passage, although they were not explicitly instructed to do so (Kosslyn & Matt, 1977). In that study, it is possible that subjects may have associated each written passage with a particular speaker and felt an expectation to reproduce the rate of that speaker.

Another aspect of speech that persists is syntax. When listeners were asked to repeat a sentence they had heard and then produce a description of a picture, they tended to use the same syntactic form as in the former sentence to describe the scene (Bock, 1986). For example, when subjects heard and repeated the sentence, “The referee was punched by one of the fans,” they were more likely to describe a picture with a church and a lightning bolt as “The church is being struck by lightning,” with both sentences in the passive form (Bock, 1986).

A few studies suggest that the tempo of music performances persists across sequences. Cathcart and Dawson (1928) instructed pianists to perform one melody at a particular tempo and another melody at a faster or slower tempo. When pianists attempted to perform the first melody again at the original tempo, their tempo drifted in the direction of the second melody. More recently, Warren (1985) reviewed studies of tasks that varied from color judgments to lifting weights. Each domain displayed a perceptual homeostasis, which Warren (1985) termed the “criterion shift rule:” that the criterion for perceptual judgments shifts in the direction of stimuli to which a person has been exposed. Warren (1985) suggested that a criterion shift serves to calibrate perceptual systems so that behavior will be appropriate for environmental conditions.

Jungers, Palmer, and Speer (2002) found evidence for temporal persistence in language and music. In a speech experiment, subjects read two sentences aloud as a measure of their preferred speech rate. Then they heard a prime sentence followed by a written target sentence matched for number of syllables, lexical stress pattern, and syntactic structure on each trial. The primes were recorded by a naive female speaker at

slow (750 ms or 80 bpm per accent) and fast (375 ms or 160 bpm per accent) rates. Subjects read the target sentences aloud. They were instructed to attend carefully to the sentences for a later recognition task. The subjects' rates were influenced by both their preferred rate and the prime rate. The music task was similar in design to the speech task, but subjects were experienced pianists. Prime and target melodies were matched for meter and length. As in the language version, both the prime rate and the preferred rate predicted the performance rate of the target melody.

The goal of the current studies is to examine whether people persist in the prosody of language and music. The prosodic persistence study by Jungers, et al. (2002) suggests that people do persist in the global prosodic dimension of tempo. This study seeks to determine if people persist in prosodic dimensions that do or do not relate to the syntactic or rule-based elements of speech and music. A series of control experiments that test listeners' syntactic and metrical interpretations serve to give base rate information for a set of sentences and melodies that listeners can clearly disambiguate, metrically and syntactically, based on prosodic cues of intensity and phrase breaks. Four experiments are reported, two of which address prosodic cues in performance and two of which address prosodic cues in perception.

The first experiment examines whether pianists persist in the intensity or the articulation of what they have just heard. The intensity patterns of strong and weak beats are tied in a meaningful, rule-based way to either a binary (2/4 or 4/4) or ternary (3/4 or 6/8) meter. Articulation is an acoustic variable that is varied across performances, but not tied to a particular meter. (Articulation is defined as the offset of one event minus the

onset of the next event, so that negative values represent staccato (separated) events and positive values represent legato (overlapping) events). Melodies heard by pianists contained either staccato or legato articulation across all events in both the binary and ternary intensity patterns. After pianists heard these melodies, they then performed melodies that were similar in number of events and musical structure. The experiment asked whether pianists persist in the rule-based (intensity-meter) or the non-rule-based (articulation) acoustic cues from what they have heard before when they perform similar melodies. Pianists were told to concentrate on the melodies for a later recognition task. The melody recognition task addressed whether intensity or articulation cues influence later memory recognition of these melodies.

If pianists form a representation that includes only categorical information (Raffman, 1993), then they would not be expected to produce or remember either the articulation cues or the intensity patterns. These cues should be used to form the representation of the melody, but then be lost. Another possibility is that pianists will only focus on and persist in the intensity cues identifying meter, since this is part of the musical structure, while the articulation cues will not be retained.

The second experiment, a language production study, examines whether speakers persist in the phrase break or the tonal pattern of what they have just heard. The phrase break location (placed early or late in the sentence by a phonetically trained speaker instructed to use particular breaks and tonal patterns) is correlated with the syntactic interpretation of the sentence. The tonal pattern (H-L% or L-H%, heard at the phrase break), another prosodic dimension, is varied across utterances, but is not tied to a

particular syntactic interpretation. Listeners heard a sentence and then produced a sentence similar in number of syllables and grammatical structure. Experiment 2 asked whether speakers persist in the syntactic (phrase break) or the non-syntactic (pitch pattern) acoustic cues. Listeners were also asked to remember the sentences for a later recognition task. The sentence recognition task addressed whether these syntactically-related or syntactically-unrelated cues influence recognition memory for the sentences.

If listeners form a reduced representation of each sentence based solely on the words of a sentence, then neither prosodic phrase break nor tonal pattern cues may persist. If listeners form a representation that includes sentence meaning, without acoustic details, they may still persist in the phrase break since it relates to the meaning, but they will not persist in the pitch pattern. There is evidence that speakers persist in the syntax of a sentence (Bock, 1996). Since the phrase break prosodic cue relates to the syntax, then participants may persist in the phrase breaks only. If listeners form a representation that includes all the acoustic details of a sentence, then they are more likely to persist in both the tonal pattern and the prosodic phrase break. This result is possible since there is evidence that speakers do remember non-syntactic prosodic details (Speer, et al., 1993). It should be noted that even if listeners perceive and remember the acoustic details, this does not mean the listeners will then produce these acoustic features in a subsequent utterance.

The last two experiments test the possibility that a second group of listeners can detect the syntax or meter that persisted in the musicians' and speakers' productions. The third experiment, a music perception experiment, examines whether listeners can

correctly interpret the intended meter of productions from the performance experiment.

The productions of four pianists were included. Two of the pianists used intensity cues to indicate meter and two of the pianists used articulation cues. Will listeners be able to identify the meter of the performances using the intensity or articulation cues? If the listeners in the music perception experiment are able to use the pianists' target performances to identify the original meter, then there is evidence that prosody can persist through more than one listener.

The fourth experiment, a speech perception experiment, examines whether listeners can interpret the syntax of the productions from four of the speakers in the speech production study. Listeners were asked to identify the intended meaning of each produced sentence and choose one of two interpretations. If listeners can accurately interpret the syntax of the productions that match the original prime sentence syntax, it suggests that persistence could be useful in a conversational context.

CHAPTER 2

CONTROL EXPERIMENT 1: MUSIC STIMULI

The goal of this experiment was to identify a set of melodies whose meter can be determined by listeners through intensity cues in staccato and legato computer-controlled performances. In order to make claims about persistence, it was important to begin with stimuli that contained salient acoustic cues that listeners clearly perceived. Thus, if a pianist's performance contained particular acoustic cues, the claim that these cues related to the original heard performance was possible. In this experiment, listeners heard metrically ambiguous melodies, performed with intensity cues indicating either a binary or ternary interpretation, as well as a control version of the melodies that contained no intensity cues (all events were the same intensity). Listeners chose "binary" or "ternary" and their confidence in their decision. The set of stimuli for which their "binary" or "ternary" decisions agreed with the intensity cue pattern in both the staccato and legato versions were to be used later in the music performance study.

Method

Participants. Twenty-six musically trained listeners participated in the study. Twenty-five participants had formal training on a musical instrument (mean yrs of private lessons = 6.78, range = 2 to 12 yrs). One participant did not have any private lesson

experience, but had performed an instrument in a band for 8 yrs and was included in the study. Participants received course credit in an introductory psychology course. None of the subjects reported having any hearing problems.

Apparatus. The musical stimuli were heard over speakers with a piano timbre generated by a Roland RD-600 keyboard and amplified through a Superscope stereo amplifier.

Materials. The stimuli consisted of 30 short, isochronous melodies that were 13 quarter-note events long. Each melody was composed to be metrically ambiguous so that the melodic contour did not clearly indicate either a binary (2/4 or 4/4) or a ternary (3/4 or 6/8) meter. See Figure 2.1 for a metrically ambiguous sample melody. The interonset interval (IOI, measured from onset-to-onset) for each quarter-note event was 500 ms. Two different articulation versions for each melody were created: staccato and legato. In the staccato version, there were 350 ms between the offset of one event and the onset of the next event; the duration of each tone was 150 ms. In the legato version, there were 10 ms between the offset of one event and the onset of the next event; the duration of each tone was 490 ms. Musically, staccato is described as a detached style and legato is described as a connected style of performance. For each articulation version, three intensity patterns were created on the computer: control, binary, and ternary. In the control version, all of the note events were the same intensity. In the binary version, the event intensities alternated between strong and weak. In the ternary version, the event

Figure 2.1. Sample metrically ambiguous melody

intensity pattern was strong, weak, weak. Each of the 30 melodies had 6 versions, based on articulation and intensity patterns: staccato-control, staccato-binary, staccato-ternary, legato-control, legato-binary, and legato-ternary.

An amateur pianist performed a subset (melodies 1-5) of the stimuli on a Roland RD-600 keyboard to determine appropriate values for the intensity levels and staccato/legato articulation values. The performer had 12 years of private piano lessons. The pianist first performed each of the 5 melodies in a binary and ternary style to a metronome set to 500 ms. The intensity was measured in MIDI units that are correlated with keystroke velocity. The average note intensity for metrically strong beats across all performances in both meters was 85. For the binary performances, the odd beats were metrically strong. For the ternary performances, every third beat, beginning with the first beat, was metrically strong. The thirteenth (final) event was not included in the analysis. The average note intensity value for metrically weak beats across both meters was 52. The average intensity value across all events in both meters was 66. These numbers became the intensity values for the accented and non-accented stimuli events. The control condition consisted of all events at the intensity level of 66.

The same pianist performed these melodies on a separate occasion to determine the appropriate duration (ISI, measured from onset-to-offset) for the staccato and legato versions of the experiment. The pianist was instructed to perform each of the five melodies in a staccato and legato style. On this occasion, the pianist did not play to a metronome, but the average IOI was 394 ms per quarter-note event. The average duration (onset-to-offset) of the notes was 392.8 in the legato version (approximately

100% of the IOI) and 125 in the staccato version (approximately 30% of the IOI). Based on these values, the staccato duration for the experimental stimuli was set to 150 ms, 30% of the 500 ms IOI and the legato version was set to 490, 98% of the 500 ms IOI.

Design. There were 6 versions of each of the 30 melodies, for a total of 180 performances created on the computer. These 180 performances were divided by articulation type into 2 sets of 90 performances: a staccato set and a legato set. Within each set, the same melody (13-note sequence) was heard 3 times, once each for the control, binary, and ternary intensity patterns. Each participant was presented with one of the two sets so that the participant heard either the staccato or the legato version of the three intensity patterns. The first independent variable, articulation, had two levels (staccato and legato), and was a between-subject factor. The second independent variable was intensity pattern with three levels (control, binary, ternary) and was a within-subject factor. The sets were arranged so that the same melody (13-note sequence) would not be heard with different intensity patterns on adjacent trials and so that no intensity pattern (control, binary, ternary) would be heard more than three times consecutively.

Procedure. Listeners heard a melody over speakers repeated twice on each trial with 1 second of silence between repetitions. and they were instructed to judge whether the melody was performed in a binary (2/4 or 4/4) or ternary (3/4 or 6/8) meter. The participants circled the word “binary” or “ternary” and also circled their confidence in their decision on a scale of 1 to 3, with 1 = not confident and 3 = confident.

Results

A one-way analysis of variance (ANOVA) on percent ternary response by intensity (control, binary, ternary) collapsed across the articulation conditions (staccato/legato), revealed significant differences among the three intensity patterns, $F(2, 50) = 179.37, p < .05$. The mean response differed in the three conditions, with participants responding correctly in the binary and ternary conditions: control = .43, binary cues = .11, ternary cues = .91. The “binary” and “ternary” choices were combined with the confidence scale to create a scale with 1 = confident binary and 6 = confident ternary, with middle values showing less confidence. An analysis of variance on the rating scale showed the same pattern of results as the binary/ternary decision, with significant differences among the intensity versions, $F(2, 50) = 157.54, p < .05$. The scale means showed the same pattern: control = 3.26, binary cues = 1.84, ternary cues = 5.35.

An ANOVA on percent ternary response for the articulation and intensity variables showed a main effect of intensity ($F(2, 48) = 178, p < .05$), but there was no main effect of articulation and no interaction of articulation and intensity. In addition, the individual melodies were examined to ensure that they each followed the expected pattern of metrical interpretation. Four melodies were not clearly perceived by subjects as binary or ternary and they were not included in the music performance experiment. Also, two melodies showed a strong binary bias in the control condition and were not included in the music performance experiment. Thus, 24 melodies remained for which unambiguous metrical judgments were given in the presence of intensity cues.

Discussion

This experiment was conducted to find a set of metrically ambiguous melodies that listeners would perceive as binary or ternary when given intensity cues that denote that meter. In order to make claims about persistence, the initial melodies must be shown to be clearly identifiable as “binary” or “ternary” by listeners. Participants successfully used the intensity patterns to make judgments about meter, but found the melodies to be metrically ambiguous in the absence of intensity cues. As expected, the articulation pattern did not influence their decision because the articulation cues were not correlated with the metrical structure.

CHAPTER 3

CONTROL EXPERIMENT 2: SPEECH STIMULI

In order to make claims about speech persistence, it was necessary to identify linguistic utterances that contain salient acoustic cues that listeners can clearly perceive. Thus, if a speaker's utterances contain particular acoustic cues, the claim that these cues relate to the original heard sentence is possible. In this experiment, listeners heard syntactically ambiguous sentences, produced with either early or late prosodic phrase breaks. For example, the sentence, "She spoke to the child with an accent" was heard with an early break after "spoke" or a late break after "child." In addition, a version of each sentence was presented that contained either H-L% or L-H% intonational pitch patterns at the phrase break. The L-H% pattern was marked by a pitch drop followed by a small pitch rise. The H-L% was marked by a steady pitch located in the higher part of the pitch range of the speaker.

Method

Participants. Sixteen adult listeners participated in the study. Most listeners had little or no formal musical training (mean = 1.5 yrs private lessons). Participants received course credit in an introductory psychology course. None of the subjects reported having any specific hearing problems.

Apparatus. The speech stimuli were presented by computer and were heard over AKG K270 headphones at a comfortable listening level.

Materials. The stimuli consisted of 54 syntactically ambiguous sentences. There were four types of ambiguous sentences. One sentence type included an adjective that ambiguously modified either a conjoined noun phrase or the first of the two nouns in this phrase, as in the sentences, “The boy bought *black shoes and socks*” and “The *old cat and dog* were the last to be adopted.” The second sentence type included an ambiguous prepositional phrase attachment, as in, “She spoke to the child *with an accent*.” Another sentence type involved conjunctions, as in, “*Pat and Shelly’s father* said it was done for now” or “Either *Brett or Mike and Kay* will come to babysit.” The final sentence type contained an ambiguous pronoun, as in, “Jay called Eric and *he* yelled at Jamie.”

Four variations of each sentence were recorded by a female phonetician who was familiar with the ToBI transcription system. The speaker was instructed to use an early prosodic break or a late prosodic break, and either a H-L% or L-H% intonational phrase boundary tone at the break. Thus, there were 4 spoken versions of each sentence: early break-H-L%, early break-L-H%, late break-H-L%, late break-L-H%. Previous work has shown that listeners expect a sentence to continue following the H-L% or L-H% intonational phrase boundary tones (Beckman & Pierrehumbert, 1986). Thus, the H-L% and L-H% intonational boundary tones can indicate similar sentence interpretations, unlike a H-H% boundary tone, which previous work has shown that listeners usually perceive as indicating a yes-no question. All sentences ended in a L-L% phrase accent and boundary tone sequence. Unlike past studies (Albritton, et al., 1996, Fox Tree & Miejer, 2000), in which the speaker read a sentence in a meaningful context or was told

to clearly disambiguate, this speaker was instructed only about the type of phrase break and pitch contour. The remaining words and breaks were not specified, but the speaker produced similar sentence types with the same prosodic pattern.

The phrase breaks were syntactically-related cues, like the intensity in the music, since the location of a phrase break was predicted to determine the syntactic interpretation. The specific tonal sequence of phrase accent and boundary tone was like the articulation in the music, because the pitch pattern that occurs at an intonation phrase break in English is not predictable from a sentence's syntactic form. There are at least six possible phrase final tonal sequences in English, including H-L%, !H-L%, H-H%, !H-H%, and L-L%, so that the presence of a phrase break does not specify the type of pitch pattern. One difference between the articulation cue in the music and the pitch pattern in the speech is the scope of the cue. The intensity (staccato or legato) was a global prosodic cue, appearing on each event. The pitch pattern (H-L%, L-H%) was a local prosodic cue, appearing at a phrase break location.

The syntactic interpretation of the sentence was related to the phrase break location, but the particular phrase break locations and possible sentence interpretations depended on the type of sentence. Table 3.1 lists the four sentence types and gives an example for each interpretation.

1. Adjective that modifies either a conjoined noun phrase or the first of the two nouns

Early prosodic phrase break: The boy bought black shoes / and socks.

Interpretation: The shoes are black, but the socks may be black.

No prosodic phrase break: The boy bought black shoes and socks.

Interpretation: Both the socks and shoes are black.

2. Prepositional phrase attachment

Early prosodic phrase break: She startled / the man with the gun.

Interpretation: The man had the gun.

Late prosodic phrase break: She startled the man / with the gun.

Interpretation: She had the gun.

3. Conjunction

Early prosodic phrase break: Either Brett / or Mike and Kay will come to babysit.

Interpretation: Brett will come alone or Mike and Kay together will come.

Late prosodic phrase break: Either Brett or Mike / and Kay will come to babysit.

Interpretation: Kay will come with one of the two men (Brett or Mike.)

4. Ambiguous pronoun

Pitch accent on “and”: Ruth hit Kate AND she hit Jason.

Interpretation: Ruth hit Jason.

Pitch accent on pronoun: Ruth hit Kate and SHE hit Jason.

Interpretation: Kate hit Jason.

/ = prosodic phrase break

CAPITAL letters = pitch accent

Table 3.1. Control Experiment 2 and Experiment 2: Four types of syntactically ambiguous sentences.

The sentence, “Either Brett or Mike and Kay will come to babysit” was produced with a phrase break after “Brett” (early break) or after “Mike” (late break). The early break, “Brett/ or Mike and Kay,” suggested that Brett alone or Mike and Kay together will come to babysit. The late break, “Brett or Mike/ and Kay,” suggested that Kay will come with one of the two men (either Brett or Mike). Sentences with an ambiguous prepositional phrase, such as, “She spoke to the child with an accent” contained either an early break after “spoke” or a late break after “child.” For these sentences, a break after “spoke” suggested that the child had the accent, a low syntactic attachment interpretation. A late break after “child” suggested that the main subject, “she” had the accent, a high syntactic attachment interpretation. The ambiguous adjective sentences did not have a true “early” and “late” break. Instead, there was an early break after the second noun, as in, “He bought black shoes/ and socks.” The “late break” version contained no breaks within the sentence.

Unlike the music stimuli, in which a computer-generated controlled intensity version is possible, it is difficult to create a neutral or control sentence. Instead, the linguistic control consisted of the same sentences presented visually (in text) on a computer screen to assess readers’ interpretation without any auditory stimulus.

Design. There were 4 spoken versions and 1 text version of each of the 54 sentences, for a total of 216 spoken utterances and 54 text sentences. The 216 spoken utterances were divided by intonational phrase boundary tone into 2 sets of 108 utterances: a H-L% set and a L-H% set. Within each set, the same sentence was heard twice, once with an early phrase break and once with a late phrase break. Each participant

was presented with one of the two sets so that the participant heard either the H-L% or the L-H% version of the two phrase breaks for each sentence. Every participant saw the same 54 text sentences. The first independent variable, boundary tone, had two levels; H-L% and L-H%, and was a between-subject factor. The second independent variable was phrase break with three levels (control (text), binary, ternary) and was a within-subject factor.

Procedure. In the first block of trials, the 54 control sentences were presented in text on the computer screen. On each trial, participants read the sentence silently and three seconds later, two possible written interpretations appeared, one on each side of the computer screen. For example, participants read, “She threatened the man with a gun.” The two interpretations were, “A) She had the gun.” and “B) The man had the gun.” Participants circled A or B on their answer sheet as well as their confidence in their decision on a scale of 1 to 3. Participants then filled out a questionnaire on their language background. This questionnaire was inserted at this point so that participants would be able to focus on the auditory stimuli and would be less focused on the written sentences they had just seen.

In the second block of trials, the early and late phrase break spoken versions of each sentence were presented in random order. On each trial, participants heard a spoken sentence over headphones. Three seconds later, two written interpretations appeared on the screen and the spoken sentence was repeated. Subjects chose interpretation A or B and their confidence in their decision. The experiment lasted approximately 50 minutes.

Results

A one-way analysis of variance on the mean percentage late break interpretations by early and late phrase breaks showed a significant difference between the phrase break conditions, $F(1,15) = 206.0$, $p < .05$, seen in the means: early break = 17.8% and late break = 85.9%. The interpretation choices were combined with the confidence scale to create a scale with 1 = confident early break and 6 = confident late break, with middle values showing less confidence. This scale showed the same pattern of results as the basic early/late decision.

A two-way ANOVA on responses by phrase breaks and boundary tones showed there was no significant difference in responses to sentences with H-L% and L-H% boundary tones. The individual sentences were examined to ensure that they each followed the expected pattern of syntactic interpretation. Based on this control experiment, twenty-four utterances were chosen that listeners correctly interpreted based on the intended phrase break cues in both the H-L% and L-H% versions.

Discussion

This experiment was conducted to identify a set of syntactically ambiguous sentences that listeners interpreted as either high or low syntactic attachment when they heard a phrase break that signified that syntactic interpretation. Listeners were influenced by the phrase break in identifying the syntactic interpretation regardless of boundary tone (H-L% and L-H%) versions.

CHAPTER 4

EXPERIMENT 1: MUSIC PRODUCTION

The goal of the music performance experiment was to determine if pianists persist in the prosody of a performance they have just heard. More specifically, do pianists persist in the metrically-related (intensity) or the metrically-unrelated (articulation) acoustic cues? Pianists heard either a strong-weak or a strong-weak-weak pattern of intensity which related to a binary and ternary interpretation. Pianists also heard staccato and legato performances. These articulation variables do not relate to any particular meter. On each trial, pianists heard a melody and then were asked to perform a different melody. The melodies were blocked by the intensity cue pattern. The pianists were not told to imitate what they had heard. Instead, the focus of the experiment was on the concluding memory test. Pianists were instructed that they would be tested on their memory for the melodies at the end of the experiment. Their performances were examined for intensity, IOI, and articulation.

Method

Participants. Sixteen adult pianists who had taken formal piano lessons (mean yrs of lessons = 9.375, range = 5-13) participated in the study. Participants received course credit in an introductory psychology course or a nominal fee for their participation. None of the subjects reported having any hearing problems.

Apparatus. Participants heard musical stimuli with a “concert grand” piano timbre generated by a Roland RD-600 keyboard over AKG headphones and they performed on a Roland RD-600 digital piano. The pitch events, timing, and keystroke velocity were recorded using FTAP (Finney, 2001).

Materials. Twenty-four melodies taken from the control experiment served as stimuli. These melodies were perceived as metrically ambiguous when heard without intensity changes, but perceived as binary or ternary when the intensity pattern of strong and weak beats indicated a meter. Only the intensity-marked versions (binary, ternary) of these melodies were included. These melodies were paired as prime and target melodies. The melodies in each prime/target pair were both major or both minor, but they were in different musical keys. The musical notation for the target melodies did not include time signature (meter) or dynamic (intensity) indications.

Design. The independent variables in this within-subject design were intensity pattern (binary, ternary) and articulation (staccato, legato). The intensity pattern was blocked so participants heard binary trials followed by ternary trials or vice versa. The intensity pattern was also counterbalanced with meter within a stimulus so that a given melody was heard in its binary intensity pattern for half of the subjects and its ternary intensity pattern for half of the subjects. The articulation was randomized across stimuli so that half of the stimuli were staccato and half were legato, but fixed for each stimulus; an individual prime melody was either staccato or legato for all participants. The member of the stimulus pair (prime, target) and the order (first half, second half) were

counterbalancing variables. The prime and target melodies alternated so a given melody was heard as a prime melody for half of the pianists and performed as a target melody for the other half.

Pianists alternated listening to and performing melodies on each of ten trials, with a block of five trials in which the prime was binary and a block of five trials in which the prime was ternary. Thus, each pianist heard a total of ten prime melodies and performed ten target melodies. In addition, each pianist began with a practice trial. The prime melody in the practice trial contained the same intensity cues as the prime in the first experimental trial.

The dependent variables were intensity, interonset interval (IOI, onset-to-onset), interstimulus interval (ISI, onset-to-offset), and adjusted articulation (ISI/IOI) across all events and on the expected metrically strong and weak beats.

A fifteen-item memory test on prime melodies only followed the prime/target portion of the experiment and included four types of items: three “same” (identical to prime melodies), six “new” (not included previously in the experiment), three “same intensity-different articulation”, and three “different intensity-same articulation”. The items in the memory test were identical to or related to those items that had been primes.

Procedure. The pianists first sight-read two melodies notated without meter indication to assess their preferred intensity and timing. Next, pianists listened to a prime melody and then performed a notated target melody on each of 11 trials. The first of these trials was a practice trial. The experimental trials were blocked in two groups of five prime-target pairs that contained the same intensity cues. Between the two blocks,

participants filled out a questionnaire about their music experience. The pianists were instructed to pay careful attention to the melodies because they would be asked to recognize them later. Following the prime and target melodies, pianists listened to a 15-item memory test. On each trial, the melody was repeated two times. The pianists answered “yes” or “no” to whether they had heard this melody before. They also circled their confidence in their decision on a scale of 1 to 5, with 1 = not very confident and 5 = very confident.

Results

The offset time, onset time, and intensity levels for each note event were measured. The 13th event was the final event in each melody. Because the timing for the final event is unspecified, this event’s intensity and offset were not included in the analysis. The IOI (interonset interval) was measured from the beginning of one event to the beginning of the next event. ISI was calculated by subtracting an event’s offset minus its onset. The adjusted articulation was calculated by dividing the ISI by the IOI, to show the percent of the IOI during which the event sounded. An analysis of variance on IOI by articulation priming condition (staccato/legato) showed no differences between articulation priming conditions, $F(1,15) = 2.19$, n.s. An ANOVA on IOI by intensity (binary/ternary) and beat (strong/weak) showed no significant effects and no interaction. This is not surprising since IOI is a representation of the tempo of the performance and the priming melodies were heard with the same IOI throughout.

An ANOVA on ISI by articulation condition showed a significant difference in ISI following a staccato or legato prime, $F(1,15) = 6.045, p < .05$. Pianists played with a more separated style following the staccato priming melodies than following the legato priming melodies. However, this analysis does not take the tempo into account and the difference may simply be due to faster performances following the staccato prime. To control for this, the ISI was divided by the IOI, forming an adjusted articulation, to show the percent of the IOI during which the event was sounded. A legato event would be 1 or greater than 1 and a staccato event would be less than 1. There was a significant difference between the adjusted articulation following a staccato prime vs. a legato prime: $F(1,15) = 5.15, p < .05$. Even when controlling for tempo variations, the pianists played a more separated style following the staccato than following the legato prime melodies. The average adjusted articulation in the preferred melodies was 1.00, showing that pianists performed in a legato style when the prime was not present. Figure 4.1 shows the mean ISI/IOI for pianists' target performances following legato and staccato prime melodies.

Figure 4.1. Experiment 1: Mean ISI/IOI for pianists' target performances following legato and staccato prime articulation

An analysis of performed intensity levels by meter of prime and expected strong and weak beats showed no significant differences between expected strong and weak beats, although there was a trend for the expected strong beats to be more intense than the expected weak beats. However, this does not mean that the pianists did not reproduce the meter of the prime melodies. An ANOVA on the adjusted articulation by expected strong and weak beats revealed a significant difference : $F(1, 15) = 9.105, p < .05$ with expected strong beats showing longer (legato) values than expected weak beats. See Figure 4.2 for the mean adjusted articulation (ISI/IOI) value for each event by prime binary and ternary melody conditions.

Memory results. A one-way ANOVA on percent “yes” response by type of memory stimuli (same meter/same articulation, new, same intensity/different articulation, same articulation/different intensity) was borderline significant, with listeners most often responding “no” to new melodies which were not heard during the performance part of the experiment: $F(3,45) = 2.73, p = .055$. The yes/no response was combined with the participants’ confidence in the decision to form a combined scale with 1 = confident no and 10 = confident yes. An ANOVA on the combined scale by the type of memory stimuli was significant, with participants responding no to new melodies better than previously heard melodies $F(3,45) = 4.65, p < .05$. Post-hoc tests did not show differences between the three types of old items (same meter/same articulation, same meter/different articulation, different meter/same articulation). Figure 4.3 shows the percent “yes” response for melodies with same or different prosodic cues.

Figure 4.2. Experiment 1: Mean ISI/IOI for each event in pianists' target performances following binary and ternary prime melodies

Figure 4.3. Experiment 1: Mean percent “yes” response for melody recognition task by cue condition

Discussion

Pianists showed persistence of the metrically-unrelated cue of articulation (staccato/legato). Pianists did not perform with the same intensity pattern they had heard, even though this metrically-related cue was blocked. However, pianists did not ignore the meter. In fact, they did persist in the meter, but they used articulation cues than intensity to produce a binary or ternary metrical interpretation. Pianists played metrically strong beats with longer ISIs than metrically weak beats. This metrical beat difference was true even when tempo was controlled, as in the adjusted articulation. This means pianists did perceive the meter and persisted in the meter, but they did not simply imitate the acoustic cues they had heard. Instead, they formed a representation of meter based on the intensity pattern and performed a following melody in the same meter using ISI differences. Thus, pianists' performances revealed persistence of metrically-related and metrically-unrelated prosodic dimensions.

Although listeners were able to discriminate new from previously heard melodies, the memory test did not show differentiation between the categories of melodies listeners had heard before. This may be due partly to a floor effect. The melodies were very difficult to remember. Each melody consisted of thirteen isochronous events with no rhythmic or tempo variations to differentiate the melodies. Pianists reported finding the memory task very difficult.

CHAPTER 5:

EXPERIMENT 2: SPEECH PRODUCTION

The goal of the language production experiment was to determine if speakers persist in the prosody of sentences they have just heard. More specifically, do speakers persist in the syntactically-related or the syntactically-unrelated acoustic cues? The syntactically-related cue was the location of the phrase break (either early or late). The syntactically-unrelated cue was the tonal pattern (either H-L% or L-H%), which did not relate to a particular syntactic interpretation in the sentences from the control experiment. On each of twelve trials, participants heard a sentence and then read aloud a second sentence that was similar in number of syllables and structure. The sentences were blocked by the location of the phrase break. The participants were not told specifically that they were to imitate what they had heard. Listeners were instructed that they would be “asked to listen to and to read short sentences.” They were also instructed to “pay careful attention to these sentences because you will be asked to recognize them later.”

Method

Participants. Thirty-two adults participated in the study. Participants had between 0 and 8 years of private lessons on a musical instrument (mean = 1.95). Participants

received course credit in an introductory psychology course. Thirty-one of the subjects reported having no hearing problems and one subject reported a slight problem, but did not specify the problem and was included.

Apparatus. Participants heard language stimuli over AKG K270 headphones and their voices were recorded to DAT using a head-mounted AKG C420 microphone. The auditory stimuli as well as the text for spoken sentences were presented on a personal computer.

Materials. The experiment consisted of twenty-six heard and produced sentences, referred to as prime and target sentences. These sentences were chosen from the language control experiment because the ambiguous sentence meaning was interpretable based on the location (or existence) of a break and the pitch pattern (L-H%, H-L%) did not correlate with the syntactic interpretation. The prime and target sentences were paired, so that each pair was made of the same type of sentence ambiguity (prepositional phrase, pronoun, conjunction, modifying adjective) with a similar number of syllables. The text for the target sentences did not include punctuation such as commas to mark expected phrasing. See Table 3.1 for an example of each of the sentence types.

Design. The independent, within-subject variables were phrase break (early break, late break) and intonational pitch pattern (H-L%, L-H%). The phrase break was blocked so that participants heard early break sentences followed by late break sentences or vice versa. The phrase break was also counterbalanced so that a given sentence was heard with an early break for half of the participants and a late break for half of the participants. The intonational pitch pattern was randomized within the phrase break

location so that half of the sentences were H-L% and half were L-H%. An individual prime sentence was either H-L% or L-H% for all participants. The counterbalancing variables were member of stimulus pair (prime, target), and order (first half, second half). The prime and target sentences alternated so a given sentence was heard as a prime sentence for half of the participants and produced as a target sentence for the other half.

The dependent variables were IOI of each key word, pitch pattern, and phrase break. The IOI of key words were measured from the onset of the word to the onset of the next word. Thus, both the word duration and the following pause were included. A key word was defined as the word before a possible phrase break. If participants break after this word, the word should have a longer IOI than if the participants do not break at this point. The pitch patterns and phrase breaks were determined through a ToBI analysis.

A twenty-item memory test on the prime sentences included five types of items: four “same” (identical to prime sentences), eight “new” (sentences not included in the experiment), four “same phrase break (syntax)-different tonal pattern (prosody),” four “different phrase break (syntax)-same tonal pattern (prosody),” and four “different phrase break (syntax)-different tonal pattern (prosody).” The new items contained some of the same words used in the original sentences so that participants could not use a single word or phrase to remember a sentence.

Procedure. The participants first read aloud three sentences presented in text on the computer screen to assess their preferred prosodic production. Next, participants listened to a prime sentence and then produced a target sentence on each of thirteen trials.

The first trial served as a practice trial. Twelve experimental trials allowed for six early break and six late break trials. Between blocks, participants completed a questionnaire on their music and language background. Only ten experimental trials were used in the music performance experiment because ambiguous melodies were more difficult to create than ambiguous sentences. The participants were instructed to pay careful attention to the sentences because they would be asked to recognize them later. Following the prime and target sentences, participants listened to a twenty-item memory test. On each trial, the sentence was heard over headphones two times. The participants answered “yes” or “no” to whether the words in the sentence were exactly the same as the words from a sentence they had heard before. They also circled their confidence in their decision on a scale of 1 to 5, with 1 = not very confident and 5 = very confident.

Results

The IOI of the word proceeding the possible phrase break in each sentence was measured from the onset of the word to the onset of the next word, so that pauses following the word were included. If participants persisted in the prosody of the sentences they just heard, they should produce the same phrase breaks as in the prime. This analysis was based on the assumption that the word prior to a break will be longer than the same word that is not followed by a phrase break (Cooper & Paccia-Cooper, 1980). This IOI analysis did not include two sentences that had an ambiguous pronoun because these sentences were hypothesized to rely on accent and not duration for syntactic interpretation. Table 3.1 lists the sentence types and gives an example of the possible prosodic phrase break locations. In the ambiguous adjective sentences, such as

“She served the cold *Sprite* and tea,” the duration of the noun following the adjective was measured. In the ambiguous grouping sentences, such as “*Mike* and Meg or Bob will come to..,” the duration of the first noun was measured, and in the ambiguous prepositional phrase sentences, such as “She yelled at the *kid* using the intercom,” the first noun following the verb was measured.

The average IOI for the words produced at an expected break was longer than the IOI for the words produced at a location without an expected break. The expected phrase break mean was 388 ms and the no phrase break mean was 373 ms: a 4% difference in IOI. However, an ANOVA on word IOI by break category (expected, not expected), was not significant.

A ToBI analysis was conducted on half of the utterances to determine the participants’ prosodic phrase breaks and tonal patterns (Beckman & Pierrehumbert, 1986). The utterances of sixteen participants (four participants from each of the four orders) were included. As mentioned in the introduction, ToBI is a prosodic transcription system that allows a transcriber to mark prosodic components, including tones and breaks. The tonal component is made of a series of several nested tonal events. A boundary tone (either H% (high) or L% (low)) is located at the end of an intonational phrase. Each intonational phrase is made up of one or more intermediate phrases, marked by a H- or L- phrase accent. Each intermediate phrase must contain at least one pitch accent, such as H* (high pitch accent), L* (low pitch accent), or L+H* (rising from low to high). The break index describes the amount of disjuncture between words and ranges from 0 (words grouped together) to 4 (full intonational phrase).

A ToBI analysis on the recordings of the original primes showed that the speaker produced the same phrase break and pitch accent patterns for each early and late break sentence within each type of sentence ambiguity. Thus, the speaker used one pattern for all early prosodic phrase break sentences and another pattern for all late prosodic phrase break sentences within each of the four sentence types. See Table 5.1 for the ToBI transcription of an early and late prosodic phrase break example from each sentence type by the speaker.

ToBI analysis on the participants' utterances in response to the prime sentences showed some phrase break patterns and pitch accents that were expected with persistence. Although the results for this ToBI analysis were not statistically significant with a chi-squared test on expected break of prime by produced break of speaker, the target sentence patterns seen for the four sentence types agreed with the expected phrase breaks and pitch patterns from the prime sentences. See Table 5.2 for the percent of target utterances that contained the pitch accent or phrase break at the locations specified by the prime utterances.

A ToBI analysis on phrase breaks or pitch accents only captures part of the prosodic information in a sentence. Another analysis examined the entire ambiguous phrase for each sentence, as shown in the speaker's transcription in Table 5.1. For this analysis, the speaker's prime data was compared to the participants' target data in terms of pitch accents, intermediate phrase breaks, and intonational phrase breaks (as in Pitrelli, et al, 1994). If the speaker produced a pitch accent and the participant also produced a pitch accent, this counted as one match. Also, if the speaker produced a phrase break and the participant also produced a phrase break, this counted as a match. (Likewise, a match occurred if the speaker did not produce a pitch accent and the listener also did not produce a pitch accent at that location). If the listener also produced the same type (H*, L*, etc.) of pitch accent, this counted as an additional match. A total of 2514 items (pitch accents, intermediate phrase breaks, intonational phrase breaks, and the tonal information at each location) were judged for matches in the 16 participants' utterances. Of the 2514 items, listeners' targets matched the speaker's primes for 1667 items, or 66.3%.

1. The boy bought black shoes / and socks.

Pph after noun (expected following early prime sentence)

Early	35%
No break	33%

2. She startled / the man / with the gun.

Pph after Verb (expected early)

Pph after Noun (expected late)

Early	46%	50%
Late	42%	58%

3. Either Brett / or Mike / and Kay will come to babysit.

Pph after N1(expected early)

Pph after N2 (expected late)

Early	58%	41%
Late	50%	58%

4. Ruth hit Kate and* she* hit Jason.

And (expected early)*

Pronoun (expected late)*

Early	33%	53%
Late	25%	75%

Table 5.2. Experiment 2: Percent of target utterances by participants following Early or Late prime sentences that contain pitch accent or phrase break at specified locations

An analysis was conducted to examine matches between speaker and listener sentences defined as the matching presence/absence of a pitch accent, matching presence/absence of an intermediate phrase break, or matching presence/absence of an intonational phrase break. The tones (H, L) at these pitch accent or phrase break locations were not analyzed. In this analysis, 1048 of 1365 items matched for the listener and speaker, 76.8%. The items were grouped into four categories for each sentence: “matched presence” (both prime and target contained a pitch accent or phrase break at this location), “mismatched presence” (prime had accent/phrase break but target did not have accent/phrase break at same location), “matched absence” (neither prime or target contained an accent/break), and “mismatched absence” (prime did not have accent/phrase break but target did have accent/phrase break). A chi-squared analysis showed a significant interaction of the speaker’s and listeners’ productions (*chi-square* (1) = 267.6, $p < 0.05$). See Table 5.3 for a summary of speakers’ and listeners’ productions. The listeners were more likely to produce a phrase break or pitch accent at the same location the speaker produced a break than at another location.

Speaker Prime Productions:			
Pitch accent/ phrase		Present:	Absent:
Listeners Target Productions at same location:	Present:	609	725
	Absent:	280	1308

Table 5.3. Experiment 2: Chi-square table for prime and target utterances

A ToBI analysis comparing the prime and target tones revealed similarities at pitch accent and intermediate phrase break locations. When the prime and target sentences contained a pitch accent or phrase break at the same location, 82.4% of the locations also matched for H or L tone. Phrase break location was related to the syntactic interpretation and 89.6% of the matching phrase break locations contained the same tone. Although pitch accent location was not related to syntax, 78.6% of the prime and target tones matched at pitch accent locations. Participants did not persist in the full L-H% or H-L% intonational phrase pitch pattern from the prime sentences when they produced target sentences. Participants rarely produced a full intonational phrase during the experiment.

Memory recognition task. An ANOVA on percent “yes” response by type of memory item ($F(4,124) = 36.56, p < .05$) showed significant differences, with subjects most often responding “no” to new items. As shown in Figure 5.1, listeners best remembered those sentences that contained the same phrase breaks (syntax) and the same tonal pattern (prosody) as in the experiment. This yes/no response was combined with confidence rating to create a 10-point scale with 1= very confident “no” and 10 = very confident “yes.” An ANOVA on combined confidence scale by memory item type was also significant, $F(4,124) = 45.39, p < .05$. These significant differences did not rely on the “new” items. An ANOVA on yes/no response by intonational pitch pattern (H-L%, L-H%) was not significant, $F(1,31) = 0.09, n.s.$ A 2x2 analysis of variance on percent “yes” responses grouping together items with same breaks (with same or different pitches) and different breaks (with same or different pitches), showed a significant difference between responses to the two break categories: $F(1,31) = 4.98, p < .05$. As shown in Figure 5.2, participants responded “yes” most often when the same break was present in the memory test: same break = .79, different break = .66. This effect is also true for the confidence rating scale: $F(1,31) = 6.11, p < .05$. Post-hoc tests revealed that the “new” items were significantly different from the other four conditions. Also, the “same syntax-same prosody” condition was significantly different from the other four conditions.

Figure 5.1. Experiment 2: Percent “yes” response for sentence recognition task by cue condition

Figure 5.2. Experiment 2: Percent “yes” response for same or different prosodic phrase breaks

Discussion

The ToBI analysis showed significant persistence effects for syntactically-related and syntactically-unrelated acoustic cues. Listeners produced the same phrasal events in their target utterances as they had heard in the prime. The analysis of IOI of words proceeding phrase breaks revealed some evidence of persistence in the lengthening of words before expected intermediate phrase breaks. Also, listeners' target utterances matched the speaker's target utterances for the presence/absence of prosodic phrase breaks and pitch accents at the same event location. In addition, when the prime and target sentences contained a pitch accent or phrase break at the same location, the tones matched.

The memory test revealed differences between syntactically-related and syntactically-unrelated cues. Listeners were more likely to say "yes" to having heard a sentence before if the current sentence contained the same phrase break as the earlier sentence. Note that the instructions in the language memory experiment were very specific. Listeners were told to response "yes" if the sentence contained the "same words as a sentence you heard before." Thus, listeners should respond "yes" to the twelve trials in which the sentence had the same words and "no" to the eight trials in which the sentence contained new words. Even with these specific instructions, listeners responded "yes" and had a higher combined confidence scale when the phrase break was the same as before. This suggests that listeners incorporated the syntactically-related cue of phrase break more strongly in memory than the syntactically-unrelated cue of pitch pattern.

CHAPTER 6

EXPERIMENT 3: MUSIC PERCEPTION

The goal of the music perception study was to examine listeners' ability to identify the metrical interpretation of performances that contained intensity or articulation cues to mark the meter. On average, pianists in the performance study persisted in the metrically-unrelated articulation of the original performances, but not in the metrically-related intensity. Instead, most performers used adjusted articulation cues (ISI/IOI) to mark the meter. In this experiment, participants listened to performances generated by four pianists from the music performance experiment; two pianists who consistently persisted in adjusted articulation cues to mark meter in their target productions and two pianists who consistently persisted in intensity cues to mark meter in their target productions. Listeners heard each performance repeated two times and indicated the meter they thought the performer intended.

Method

Participants. Sixteen adult (mean age = 19.125) musicians, who had taken formal lessons on an instrument (mean yrs of lessons = 6.125, range = 3-11), participated in the study. Participants received course credit in an introductory psychology course for their participation. All participants were native speakers of English. None of the subjects reported having any hearing problems.

Apparatus. Participants heard stimuli over AKG K270 headphones while seated at a computer. FTAP was used to control the stimulus presentation (Finney, 2001).

Materials. A subset of the performances of pianists in the music performance experiment was included. Experiment 1 showed that pianists used articulation differences more often than intensity differences to mark meter. In this experiment, four pianists' performances were included: two pianists who used intensity to mark metrical downbeats and two pianists who used articulation to mark metrical downbeats. The performances of two pianists who showed stronger intensities on the expected strong beats (mean = 70.58) than on the expected weak beats (mean = 52.45) were included as the intensity performances). These pianists' performances did not show large differences in adjusted articulation on strong beats and weak beats. The performances of two other pianists who showed longer adjusted articulation on strong beats (mean = .97) than weak beats (mean = .94) but little intensity differences were included as articulation performances.

Design. Thirty-three error-free performances by the four pianists were arranged so that no more than 3 binary or ternary performances were in a row and no single pianists' performances were in a row. There were 17 binary trials (10 articulation / 7 intensity) and 16 ternary trials (7 articulation / 9 intensity). Seven of the 40 experimental target trial performances contained errors and were not included. Listeners heard one performance repeated twice on each of 33 trials in this within-subject design. The independent variables were cue type (intensity/articulation) and meter (binary/ternary). The dependent variables were metrical response (binary/ternary) and confidence rating.

Procedure. Listeners sat in front of a computer and listened to the melodies over AKG K270 headphones. They were instructed to choose the meter they thought the performer intended by circling “binary” for 2/4 or 4/4 melodies and “ternary” for 3/4 or 6/8 melodies. Each melody was heard two times. Listeners also circled their confidence in their decision, with 1 = not very confident and 5 = very confident. Listeners were asked informally if they were familiar with the distinction between the meters (2/4, 4/4, 3/4, 6/8). All subjects knew the distinction, although some subjects had not used the terms “binary” or “ternary” before participating in the experiment.

Results

A two-way ANOVA on percent ternary response by primed meter and cue type (intonation / articulation) showed no main effect of meter (binary/ternary). However, there was an interaction of primed meter and cue type: $F(1,15) = 15.5, p < .05$. As shown in Figure 6.1, intensity cues led to more accurate identification of primed meter while articulation cues led to inaccurate identification of primed meter. The response (binary/ternary was combined with the confidence rating to form a 10-point confidence scale). A two-way ANOVA on this combined scale for response and confidence by primed meter and cue type showed the same interaction between cue type and primed meter: $F(1,15) = 17.9, p < .05$.

Figure 6.1. Experiment 3: Percent “ternary” response in music perception task with intensity or articulation cues

In addition, the confidence scale showed a main effect of cue type, with higher values for intensity than articulation: $F(1,15) = 5.6, p < .05$. This increased scale response can be attributed partly to the overall increased confidence on intensity trials, regardless of the meter. An ANOVA on confidence alone showed higher confidence for intensity (mean = 3.68) than articulation (mean = 3.42) trials: $F(1,15) = 6.38, p < .05$, and higher confidence for ternary (mean = 3.63) than binary (mean = 3.47) trials: $F(1,15) = 5.25, p < .05$.

Discussion

Listeners were more accurate at identifying the primed meter when the performance contained metrically-related intensity cues than when the performance contained metrically-related articulation cues. Listeners identified the meter of the target melody through intensity cues that persisted from the prime melody.

A single experiment alone is not sufficient to draw conclusions about persistence in an ensemble context. This experiment included only some of the performances from the music performance study. However, it demonstrates that it is possible for listeners to interpret the meter when the target performance contains the same acoustic details as the prime performance. One concern is the perceptual salience of the two cues, intensity and articulation. Although the performances were chosen so that a distinction between articulation and intensity cues could be drawn, the intensity cues may have more strongly indicated the meter than the articulation cues in this set of performances.

A more natural context, in which instrumentalists are trading a melody back and forth may reveal stronger effects of this continuing persistence. The listeners in the laboratory often tapped their feet or moved their bodies to the performances in order to determine the meter. In a natural setting, this bodily incorporation of the meter would be more pronounced. Also, it is possible that listeners could use both the acoustic cues as well as visual cues when determining and persisting in the meter of a live performance. At the very least, this experiment suggests that the theme of prosodic persistence in an ensemble setting may be worth pursuing.

CHAPTER 7

EXPERIMENT 4: SPEECH PERCEPTION

The goal of the speech perception study was to examine the phenomenon of persistence in a situation that is closer to natural spoken conversation. Listeners judged the interpretation of sentence that contained prosodic phrase break acoustic cues. If listeners in the speech production study persisted in the acoustic cues of the sentences they heard, could other listeners hear their utterances and identify the syntactic interpretation of the original sentences? Listeners heard the productions of four speakers from the speech production experiment and judged the speakers' intended syntactic interpretation.

Method

Participants. The same sixteen adults who participated in the music perception study also participated in this experiment. All participants were native speakers of English.

Apparatus. Participants heard stimuli over AKG K270 headphones while seated at computer.

Materials. A subset of the productions from the language production experiment was included. Forty-four error-free productions of four speakers were included. (4 of the 48 experimental utterances contained errors such as unnatural pauses or mispronounced

words). The particular speakers were chosen because their ISI for the word before a prosodic phrase break matched the syntactically-related phrase break cues from the prime for six or more of their ten utterances.

Design. Forty-four error-free performances by the four speakers were arranged so that no more than 3 early or late primed productions were in a row and no single speakers' productions were in a row. There were four types of sentences: 15 adjective that modifies either a conjoined noun phrase or the first of the two nouns, 11 prepositional phrase attachment, 11 conjunction, and 7 ambiguous pronoun. Listeners heard one performance repeated twice on each of forty-four trials. The independent variable was primed break condition (early break/late break). The dependent variables were response (early/late break interpretation) and confidence rating.

Procedure. Participants were instructed to choose the sentence interpretation they thought the speaker intended. On each trial, they heard an utterance, saw two possible sentence interpretations on the computer screen, and then heard the sentence again. The sentence interpretations were labeled A and B and participants circled A or B on an answer sheet. Participants also circled their confidence in their decision with 1 = not very confident and 5 = very confident. The 'A' interpretation appeared on the left side of the screen and the 'B' interpretation appeared on the right side. For half of the participants, 'A' was the early phrase break interpretation and for half of the participants, 'B' was the early phrase break interpretation.

Results

An ANOVA on the percentage late break interpretation responses by phrase break (early/late prime) showed significant differences between the expected early and late interpretations: $F(1,15) = 17.35, p < .01$. Listeners responded with the late interpretation on 72.4% of primed late break trials and with the early interpretation on 40.9% of primed early break trials. An analysis by item showed participants interpreted 61.4% of the stimuli in the expected (primed) direction, which was significantly different from chance ($t(1, 42) = 5.04, p < .05$).

Discussion

Listeners responded differently to the early and late phrase break sentence trials. Although overall accuracy was not high, listeners interpreted the late break sentences in the direction of the original syntactically-related cues. Thus, the persistence effects seen in Experiment 2 were sufficient to influence other listeners' syntactic interpretations. Listeners had a bias to respond with the late break interpretation. They responded with the late prosodic phrase break interpretation 65.8% of the time. Despite this bias, there was still a difference between listeners' response to the early and late prosodic phrase break trials. Note that the labels for early and late prosodic phrase break refer to the prime sentences and not to the acoustic properties of the target sentences heard by the listeners. The assumption is that if the speakers in the sentence production experiment persisted in the syntactically-related prosodic phrase breaks, then the listeners in this speech perception study would be able to determine the phrase breaks from the target

sentences. As in the music perception experiment, this single experiment does not explain persistence in a conversational context. However, this experiment does suggest that further study in a more realistic conversation context is needed.

CHAPTER 8

GENERAL DISCUSSION

Several experiments demonstrated persistence effects for syntactically-related and syntactically-unrelated prosodic variations in the domains of music and language. In the music performance study, pianists listened to prime melodies and produced similar target melodies. The prime melodies contained metrically-related intensity cues (supports binary or ternary meter) as well as metrically-unrelated articulation cues (staccato or legato). On each trial, pianists heard a melody and then were asked to perform a different melody. The melodies were blocked by the intensity cue pattern. The pianists were not told to imitate what they had heard. Instead, the focus of the experiment was on the concluding memory test. Pianists were instructed that they would be tested on their memory for the melodies from the experiment.

Pianists performed the musical pieces with the same articulation (staccato or legato) as the performances they had just heard. These articulation cues altered from trial to trial (with each piece), so this demonstrated pianists' ability to persist in structurally-unrelated musical cues. However, pianists did not directly recreate the metrically-related intensity pattern of the prime pieces in their performances. Pianists persisted in the strong and weak beats of the prime meter, but they used timing variation instead of intensity variation. The primes contained louder intensities on the strong beats than the

weak beats and pianists performed with longer adjusted articulations (ISI/IOI) on the strong beats than the weak beats. This suggests that pianists used the intensity cues to form a representation of the meter. When the pianists performed subsequent pieces, they used the same meter as the performance they just heard, but they instantiated it with articulation cues instead of intensity cues.

Finally, pianists better remembered melodies they had heard earlier than new melodies in a melody recognition task. Pianists listened to each melody and judged whether they had heard it before. There was no difference in the pianists' ability to recognize performances with different metrically-related or metrically-unrelated prosodic cues from the original performances they heard during the experiment. The inability to differentiate performances was likely due to the pianists' overall difficulty in remembering the metrically ambiguous, isochronous melodies.

In the speech production experiment, listeners heard prime sentences and produced similar target sentences. The prime sentences contained syntactically-related prosodic phrase breaks (early or late) as well as syntactically-unrelated tonal pattern cues (H-L% or L-H%). On each trial, listeners heard a sentence and then read aloud a different sentence. The sentences were blocked by prosodic phrase break. The participants were instructed to attend to the sentences for a later sentence recognition test.

Speakers persisted in the syntactically-related phrase breaks and the syntactically-unrelated tones from the primes. Speakers were more likely to produce pitch accents and prosodic phrase breaks in the same location of their target sentences as where the pitch accents and prosodic phrase breaks occurred in the prime sentences. Also, words before

a prime phrase break were longer in IOI than the same words without a prime phrase break. Also, listeners and speakers produced the same tone (H, L) at matching phrase break or pitch accent locations.

Finally, listeners better remembered sentences they had heard before than new sentences in a sentence recognition task. In addition, listeners better remembered sentences that contained the same prosodic phrase breaks they had heard before than sentences that contained the same tonal patterns they had heard before. Thus, listeners better remembered the sentences when the syntax (indicated by prosodic phrase break) stayed the same from the experiment to the memory recognition task.

The third and fourth experiments examined listeners' ability to determine the meter or syntax of performances or utterances from the production studies. Musicians who listened to target melody performances from the music production experiment could recognize the meter when metrically-related intensity cues (but not articulation cues) persisted in the performance. Similarly, listeners who heard target sentences from the speech production experiment were better than chance at choosing the phrase structure interpretation that matched the original prime sentence. These perception experiments suggest that it is possible for listeners to determine the meter and syntax from prosodic cues that persisted in the productions.

One limitation of the current study is the naturalness of the task, particularly for the speakers. Listeners heard sentences with strong prosodic patterns (both phrase break and tonal patterns) and produced speech by reading sentences. More natural productions and more prosodic variability may be possible if participants do not read the sentences,

but instead, produce sentences from memory. More natural context situations, such as conversation during a game, show greater prosodic variation than read speech (Schafer, Speer, Warren, & White, 2000). Future studies of prosodic persistence should involve more natural contexts, so that the speech contains the full range of prosodic variability. Also, the number of stimuli in the speech study was relatively limited. More stimuli or a prolonged exposure to a break pattern could lead to greater persistence.

Prosodic persistence

The experiments suggest that prosodic persistence is the continuation of acoustic variations (both structurally-related and structurally-unrelated) from perception to production. In addition, this persistence may be involved in conversational speech or ensemble performances, as demonstrated by listeners' ability to hear target productions and recognize the primed syntax or meter. In both the speech and music production experiments, participants incorporated prime acoustic details into their own target productions. This suggests that listeners formed representations of the melodies or sentences that included more than just the notes or words.

Did listeners persist in information that was not categorical? According to Raffman (1993), listeners may forget the prosodic cues once they are used to form categories of pitch or duration. This approach does not explain pianists' ability to perform with different articulation cues following the staccato and legato primes. These articulation cues should not be remembered because they are sub-categorical. If these articulation cues are not remembered, there is no reason for them to influence and persist

in future performances. Thus, the articulation cue persistence suggests pianists do persist in sub-categorical acoustic details. Also, this approach does not explain pianists' ability to persist in the meter, which they heard instantiated with intensity cues.

Prosodic persistence is not simple imitation: the pianists persisted in the prime meter, but they marked the meter with adjusted articulation cues in their performances instead of the primed intensity cues. Even though pianists heard strong beats marked with higher intensities than weak beats, they did not reproduce the meter with intensity cues. Pianists performed with shorter adjusted articulations across events after hearing staccato primes than after hearing legato primes. However, the pianists never performed with the same event duration from the prime melodies. In the prime staccato melodies, the duration of each note event was 150 ms, 30% of the IOI. In the target melodies following the staccato prime, the event durations were 87.6% of the IOI. In the speech production experiment, the speakers produced sentences with pitch accents and prosodic phrase breaks in the same locations as where the pitch accents and breaks were located in the prime sentences, but they often used different tones.

Finally, prosodic persistence does not exist solely to support syntactic persistence. Bock (1996) showed evidence for syntactic persistence in which participants described a new scene using the same syntax they had just heard. In the current study, listeners persisted in the prosodic cues, using acoustic cues from what they have heard. For music, the pianists' performances included both metrically-related and metrically-unrelated acoustic cues. If the pianists were to only persist in metrically-related cues, than they would not have persisted in the prime articulation cues. In speech, speakers persisted in

the pitch accent and prosodic phrase break locations from the prime sentences. In addition, they matched the tone at these locations. Although phrase break location and phrasal tone related to syntax, the matching pitch accent location and pitch accent tone related to prosody.

An alternative explanation for the persistence effect is that the first sentence implicitly primes the syntactic structure that is then regenerated in the produced target sentence (Potter & Lombardi, 1998). This explanation assumes the prime representation does not include surface structure features (Potter & Lombardi, 1998). This explanation is plausible for the syntactically-related persistence, but it does not make sense for the syntactically-unrelated persistence. If the representation of the prime does not contain surface level features, then acoustic details unrelated to syntax should not persist because they will not be regenerated with the syntax. The musicians and speakers persisted in the articulation (staccato/legato) and the prosodic tones (H/L) from the prime, suggesting that the representation included both syntactic and prosodic features.

Speech/music differences

There is evidence for syntactically-related and syntactically-unrelated persistence for both speech and music, but there were differences in degree in the two domains. The musicians persisted in the structurally-unrelated articulation (staccato/legato) and used these same cues to produce meter, but speakers showed strong syntactically-related persistence and did not produce full intonational phrase tones. This difference may be related to the different goals in the two domains. In speech, the goal is to communicate

an idea and in music, the goal is the expression of the piece. A change of syntax in speech changes the meaning of the utterance, but in music, a change of *either* meter or articulation changes the expression of the musical performance.

Further evidence of the importance of syntactically-related cues was demonstrated in the speech memory test. Listeners better recognized sentences they had heard before if the phrase break (a syntactically-related cue) was the same than if it was different, but listeners did not differentiate performances that contained a different tonal pattern from performances with the same pattern they heard during the experiment. This result is surprising since past research showed evidence for memory for non-structural prosodic cues. For example, listeners can use information such as talker identity or speech rate to recognize previously heard words (Bradlow, et al., 1999, Nygaard, et al., 1995) and listeners' memory for talkers' voices aids identification of novel words in noise (Nygaard & Pisoni, 1998). The pianists did not differentiate different cue conditions in the memory test. However, the difficulty of the task suggests that pianists' results may be due partly to floor effects.

Another possible reason for different outcomes in the speech and music experiments is that the salience of the structurally-unrelated prosodic cues may have differed in the two domains. It is possible that the tonal patterns in the speech were less salient details than the syntactic structure. The sentences were blocked by syntactic structure. Perhaps if the sentences were blocked by prosodic tones, these tones would persist to a stronger degree. Different syntactically-unrelated cues, such as style of

speaking (ex. clearly enunciate speech vs. normal speech) or location of utterance within pitch range (high or low), may persist. Also, the use of sentences with contrastive stress may reveal stronger non-syntactic persistence.

Why were cues different for production and perception in music? In music, pianists persisted in the articulation cues across events and played strong metrical beats with longer articulation than weak metrical beats. Although there was a trend, there were no significant differences in intensities for strong and weak beats. However, in the music perception task, intensity cues better indicated the meter for listeners than articulation cues. This difference between production and perception is likely due to the salience of the intensity and articulation cues in the four performances including in the perception study. By including performances with metrically-related intensity cues but not metrically-related articulation cues (and vice versa), those performances with both types of metrically-related cues were eliminated. Articulation and intensity cues may interact in performance to indicate meter. This result is slightly different than the results of Sloboda (1985), in which listeners could use articulation to determine the metrical context of a performance. Intensity was also a useful cue to meter, but only some performers used this cue (Sloboda, 1985). Thus, the ability to use prosodic cues to determine meter may depend largely on the specific performance. The pianists in the Sloboda (1985) study were more experienced than the pianists in the present performance study and they may have been more consistent in their performance expression of articulation.

Why would prosodic persistence be useful? Prosodic variations in music, whether they relate to the meter or not, are part of the expression of music. Musicians in an ensemble such as a band or orchestra coordinate their performances so that they play with the same timing and the same expression. Prosodic variations in speech help speakers disambiguate the syntax in ambiguous sentences. Perhaps persisting in the same prosody also helps speakers in a conversation communicate more clearly and receive the message more easily because they eliminate some acoustic variability when they persist in each others' prosody. The perception studies showed evidence for both speech and music; a second generation of listeners identified the structural interpretation of the original primed source through the target production. This hints at the importance of prosody for conversational and musical exchanges and suggests that further exploration of mutual priming of prosodic cues in conversational and ensemble contexts may be fruitful.

LIST OF REFERENCES

- Allbritton, D.W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 714-735.
- Beckman, M.E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11, 17-67.
- Beckman, M.E. & Elam, G.A. (1997). *Guidelines for ToBI labeling. (Version 3)*. Columbus, OH: Ohio State University.
- Beckman, M.E. & Pierrehumbert, J.B. (1986). Intonational structure in English and Japanese. *Phonology Yearbook*, 3, 255-310.
- Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355-387.
- Bock, K. (2002). Persistent structural priming: Transient activation or implicit learning? Paper presented at CUNY Sentence Processing Conference, New York.
- Bock, K. & Griffin, Z.M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129, 177-192.
- Boltz, M.G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics*, 60, 1357-1373.
- 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, 206-219.
- Cathcart, E.P. & Dawson, S. (1928). Persistence: A characteristic of remembering. *British Journal of Psychology*, 18, 262-275.
- Clark, H.H. (1996). *Using language*. NY: Cambridge University Press.
- Clark, H.H. (2002). Speaking in time. *Speech Communication*, 36, 5-13.

- Clynes, M., & Walker, J. (1986). Music as time's measure. *Music Perception*, 4, 85-119.
- Collier, G.L., & Collier, J.L. (1994). An exploration of the use of tempo in jazz. *Music Perception*, 11, 219-242.
- Cooper, A.M. & Whalen, D.H., & Fowler, C.A. (1986). P-centers are unaffected by phonetic categorization. *Perception and Psychophysics*, 39, 187-196.
- Cooper, G., & Meyer, L.B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.
- Cooper, W.E., & Eady, S.J. (1986). Metrical phonology in speech production. *Journal of Memory and Language*, 25, 369-384.
- Cooper, W. & Paccia-Cooper, J. (1980). *Syntax and Speech*. Harvard University Press: Cambridge, Massachusetts.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141-201.
- Dowling, W.J. & Harwood, D.L. (1986). *Music cognition*. Orlando: Academic Press.
- Drake, C. (1993). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society*, 31, 107-110.
- Drake, C., & Botte, M.C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, 54, 277-286.
- Drake, C., Jones, M.R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, 77, 251-288.
- Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception*, 10, 343-378.
- Ellis, M.C. (1991). Thresholds for detecting tempo change. *Psychology of Music*, 19, 164-169.
- Essens, P.J., & Povel, D.-J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, 37, 1-7.

Finney, S. (2001). FTAP: A Linux-based program for tapping and music experiments. *Behaviour Research Methods, Instruments, and Computers*, 33, 63-72.

Fox Tree, J.E. (2000). Coordinating spontaneous talk. In L. Wheeldon (Ed), *Aspects of language production*. (pp. 375-406). Philadelphia: Psychology Press.

Fox Tree, J.E., & Meijer, P. (2000). Untrained speakers' use of prosody in syntactic disambiguation and listeners' interpretations. *Psychological Research*, 63, 1-13.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed), *The psychology of music* (pp.149-180). New York: Academic Press.

Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A Major (K331): A comparison of five performances. In A. Gabrielsson (Ed), *Action and perception in rhythm and music* (pp. 81-104). Stockholm: Royal Swedish Academy of Music.

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

Grosjean, F.H., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11, 58-81.

Henderson, M.T. (1936). Rhythmic organization in artistic piano performance. *University of Iowa Studies in the Psychology of Music*, 41936, 281-305.

Jones, M.R. (1987). Perspectives on musical time. In A. Gabrielsson (Ed), *Action and perception in rhythm and music*, pp. 153-176. Stockholm: Royal Swedish Academy of Music.

Jones, M.R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, 83, 323-355.

Jungers, M.K., & Palmer, C. (2000). Episodic memory for music performance. *Abstracts of the Psychonomic Society*, 5, 105.

Jungers, M.K., Palmer, C., & Speer, S.R. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing*, 2, 21-35.

Kemler Nelson, D.G., Jusczyk, P.W., Mandel, D.R., Myers, J., Turk, A., and Gerken, L.A. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, 18, 111-116.

- Kjelgaard, M.M., & Speer, S.R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40, 153-194.
- Kosslyn, S.M., & Matt, A.M. (1977). If you speak slowly, do people read your prose slowly? Person-particular speech recoding during reading. *Bulletin of the Psychonomic Society*, 9, 250-252.
- Large, E.W., & Jones, M.R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106, 119-159.
- Large, E.W., & Palmer, C. (2001). Perceiving temporal regularity in music. *Cognitive Science*, 26, 1-37.
- Large, E.W., Palmer, C., & Pollack, J.B. (1995). Reduced memory representations for music. *Cognitive Science*, 19, 53-96.
- LeBlanc, A., Colman, J., McCrary, J., Sherrill, C., & Malin, S. (1988). Tempo preferences of different age music listeners. *Journal of Research in Music Education*, 36, 156-168.
- Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7, 106-122.
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5, 253-263.
- Lehiste, I., Olive, J.P., & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, 60, 1199-1202.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- Levelt, W.J.M., (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levitin, D.J., & Cook, P.R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, 58, 927-935.
- Martin, J.G. (1970). Rhythm-induced judgments of word stress in sentences. *Journal of Verbal Learning and Verbal Behavior*, 9, 627-633.
- Merker, B. (2000). Synchronous chorusing and human origins. In N.L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music*. (pp. 315-327). Cambridge: MIT Press.

Miller, J.L., Grosjean, F., and Lomato, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, 41, 215-225.

Munhall, K., Fowler, C.A., Hawkins, S., & Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics*, 20, 225-239.

Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3), 355-376.

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1993). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics*, 57, 989-1001.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 331-346.

Palmer, C. (1996a). Anatomy of a performance: sources of musical expression. *Music Perception*, 13, 433-454.

Palmer, C. (1996b). On the assignment of structure in music performance. *Music Perception*, 14, 21-54.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48, 115-138.

Palmer, C., Jungers, M.K., & Jusczyk, P.W. (2001). Memory for musical prosody. *Journal of Memory and Language*, 45, 526-545.

Palmer, C. & Krumhansl, C.L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 728-741.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonation contours in the interpretation of discourse. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds), *Intentions in communication* (pp. 271-311). Cambridge, MA: MIT Press.

Pitrelli, J., Beckman, M.E., & Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of the 1994 International Conference on Spoken Language Processing* (pp. 123-126). Yokohama, Japan.

Pisoni, D.B. (1997). Some thoughts on “normalization” in speech perception. In K. Johnson & J.W. Mullenix (Eds.), *Talker variability in speech processing*, pp. 9-32, San Diego: Academic Press.

Potter, M.C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38, 265-282.

Povel, D.J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception & Performance*, 7, 3-18.

Price, P., Ostendorf, M., Shattuck-Hufnagle, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90, 723-735.

Raffman, D. (1993). *Language, music, and mind*. (Bradford Book, ed.). Cambridge, MA: MIT Press.

Repp, B.H. (1992). Probing the cognitive representation of musical time: structural constraints on the perception of timing perturbations. *Cognition*, 44, 241-281.

Repp, B.H. (1994). On determining the basic tempo of an expressive music performance. *Psychology of Music*, 22, 157-167.

Schafer, A.J., Speer, S. R., Warren, P., & White, D.S. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29, 169-182.

Seashore, C.E. (ed. 1936). *Objective analysis of musical performance, Vol. 4*. Iowa City: University of Iowa Press.

Siegel, J.A., & Siegel, W. (1977). Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception and Psychophysics*, 21, 399-407.

Sloboda, J.A. (1983). The communication of music metre in piano performance. *Quarterly Journal of Experimental Psychology*, 35, 377-395.

Sloboda, J.A. (1985). Expressive skill in two pianists: metrical communication in real and simulated performances. *Canadian Journal of Psychology*, 39, 273-293.

Speer, S.R., Crowder, R.G., & Thomas, L.M. (1993). Prosodic structure and sentence recognition. *Journal of Memory and Language*, 32, 336-358.

Stein, E. (1989). *Form and performance*. New York: Limelight.

- Streeter, L. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*, 64, 1582-1592.
- Sundberg, J. (1993). How can music be expressive? *Speech Communication*, 13, 239-253.
- 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.
- Wales, R. & Toner, J. (1979). Intonation and ambiguity. In W.E. Cooper and E.C.T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, N.J.: Erlbaum.
- Warren, P. (1999). Prosody and sentence processing. In S. Garrod and M. Pickering (Eds.), *Language processing*. (pp. 155-188). Hove: Psychology Press.
- Warren, R.M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review*, 92, 574-584.
- Wayland, S.C., Miller, J.L., & Volaitis, L.E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America*, 95, 2694-2701.
- Windsor, W. L., & Clarke, E. F. (1997). Expressive timing and dynamics in real and artificial music performances: Using an algorithm as an analytical tool. *Music Perception*, 15, 127-152.
- Wingfield, A., & Klein, J.F. (1971). Syntactic structure and acoustic pattern in speech perception. *Perception and Psychophysics*, 9, 23-25.