ABSTRACT

CUES TO WORD SEGMENTATION FOR ADULT LEARNERS

By Rebecca A. Horn

An important step in acquiring a language is the ability to segment words from speech streams. Typical speech contains many cues to word segmentation, but cues are not always consistent. In studying the efficacy of particular cues, it has been suggested that some non-linguistic information, such as music, may actually help with word segmentation. Although it is traditionally accepted that music and language are treated as separate types of information by the brain, recent evidence suggests that there may be shared structural, though likely not semantic, properties.

The current study was designed to compare the effects of cues to word segmentation on learning rates in order to determine if tonal information could provide a benefit beyond that provided by regular speech cues. Participants listened to a speech stream of pseudo-randomly repeated nonsense words. Speech streams were of four types: monotone, prosody-enhanced (final vowel lengthened), tonally-enhanced (each syllable 'sung' on a particular tone), and tonal-word (every 'word' 'sung' in the same series of three tones). On a forced-choice test participants were asked to choose which in a pair of syllable strings most resembled a word from the exposure stream. Learning was measured by the number of correct responses on the forced-choice test.

Results showed a significant facilitory effect of the prosodic cue (i.e., final vowel lengthening), but no effect of either tonal condition, suggesting a privileged status for language-specific cues to word segmentation. Failure to replicate previous findings of tonal facilitation are discussed in relation to the detrimental effects of two unexpectedly high between-word transitional probabilities as well as a potential lack of statistical power.

CUES TO WORD SEGMENTATION FOR ADULT LEARNERS

by

Rebecca A. Horn

A Thesis Submitted
In Partial Fulfillment of the Requirements
For the Degree of

Master of Science Psychology

Experimental

at

The University of Wisconsin Oshkosh
Oshkosh, WI 54901

May 2009

COMMITTEE APPROVAL

Frances Rauscher Advisor

6/1/09 Date Approved

James E. Ward Member

6/1/09 Date Approved

Member

6-1-09 Date Approved

INTERIM PROVOST
AND VICE CHANCELLOR

E Alan Hartman

6/4/2009

Date Approved

FORMAT APPROVAL

Gloria Splittgerber

4/24/2009

Date Approved

TABLE OF CONTENTS

INTRODUCTION

Humans begin life with the basic perceptual abilities needed for language acquisition (Kuhl, 2004) and with an innate drive to establish communication (Jusczyk, 1993). Ushakova (2000) hypothesized that vocalization is just a part, albeit an important one, of an inborn mechanism of the brain that is motivated to externalize internal states through movement. Communication is by definition a social behavior. It is thus no surprise that social interaction guides the development of language skills by providing external cues about the form personal utterances should take in order to serve a communicative function (Kuhl, 2004), and without such interaction children may not develop language at all (Johnson & Newport, 1989). Further emphasizing the idea that language acquisition is socially dependent, it has been observed that infants not only show preference for biological over non-biological sounds (Vouloumanos & Werker, 2004; Vouloumanos & Werker, 2007), but also prefer biological gestures with communicative purpose, such as American Sign Language, over pantomime gestures (Krentz & Corina, 2008). Communicative purpose can be inferred from repetitive patterns in the signal, be it vocal or physical. Thus it is appropriate that infants innately prefer periodically occurring sounds, a phenomenon known as the periodicity bias (Cutler & Mehler, 1993), and infant-directed speech, or motherese (Fernald, 1985), which may facilitate early language learning in infants by providing more consistent cue patterns than normal speech (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989).

The general drive of infants to pay attention to repetitive and purposeful sounds from other people is refined with experience to concentrate on those cues most relevant in the ambient language so that processing can become more efficient (Jusczyk, 2002). By 5 months old, infants can distinguish their own language from others (Jusczyk, 2002), and by 12 months show declines in the ability to detect phonetic differences between languages (Werker & Tees, 1984; Werker & LaLonde, 1988). These declines are a direct result of experience with language and may follow a linear trajectory with age, as opposed to the sharp cut-off suggested by the term critical period (Johnson & Newport, 1989). Decline with experience is likely due to neural commitment to the processing of relevant stimuli (Jusczyk, 1993). This developmental trend from generalized ability to specialized experience-dependant knowledge will be emphasized in all domains of language learning discussed hereafter. The current study will contribute to the literature concerning new language acquisition past the typical period for language learning by looking at cues that may assist adult learners in processing unfamiliar language input.

Statistical Learning

Because communicative gestures are pattern-based and repetitive, one can perform statistical calculations on the relationships of elements and use these to sort input into meaningful units. Research suggests that statistical learning is not domain specific (Yang, 2004), but is likely constrained by innate biases towards particular kinds of regularities (Aslin, Saffran, & Newport, 1999). Kirkham, Slemmer, and Johnson (2002) found that infants could learn statistically in the visual domain, while Saffran and Thiessen (2003) found that infants could learn statistically reliable auditory (nonlinguistic) information while inconsistent information was harder to learn. Saffran

(2002) also found participants could learn statistically in both the auditory and visual domains, but that effects were more pronounced for auditory information. This suggests both generality for statistical learning mechanisms and modality-specificity for level of effect. Specific types of input in the same modality can also show different levels of effect, but rule-learning for one type of input may facilitate rule-learning for another type of input. Illustrating this, Marcus, Fernandes, and Johnson (2007) found that exposure to language can facilitate statistical rule learning with other sounds, such as tones and animal sounds.

The type of statistical learning most discussed in the area of language acquisition is the *transitional probability*[1] (TP). A transitional probability is the likelihood that one unit will follow another. The greater the probability that a particular unit will follow another, the more likely it is that the two units are part of a cohesive segment or word. For example, for the words "baby bottle," the syllable "by" is far more likely to follow the syllable "ba" than to precede the syllable "bo" because "ba" and "by" form a word, while "by" and "bo" do not. The probability that Y will follow X is calculated by dividing the frequency of their co-occurrence (XY) by the overall frequency of Y (Brent & Cartwright, 1996; Saffran, Newport, & Aslin, 1996).

Research has shown that transitional probabilities alone are sufficient for segmenting words from a speech stream for both infants (Aslin, Saffran, & Newport, 1998; Graf Estes, Alibali, & Saffran, 2007; Saffran & Wilson, 2003; Thiessen & Saffran, 2003; see Appendix B for more information on infant research) and adults (Bonatti, Peña,

[1] Definitions of italicized words can be found in the Glossary (Appendix A).

Nespor, & Mehler, 2005; Mirman, Magnuson, Graf Estes, & Dixon, 2008; Saffran et al., 1996; Saffran, Newport, Aslin, Tunick, & Barreuco, 1997; Schön, Boyer, Moreno, Besson, Peretz, & Kolinsky, 2008; Toro, Sinnett, & Soto-Faraco, 2005), and can be used to learn higher levels of language organization such as grammar and sentence structure (Gerken, Wilson, & Lewis, 2005; Morton & Long, 1976; Saffran, 2001a; Saffran, 2002; Saffran & Wilson, 2003; Thompson & Newport, 2007). Other researchers have also suggested that learners can employ even more complex algebraic computations to learn language (Marcus, Vijayan, Rao, & Vishton, 1999).

Harris (1955) was one of the earliest theorists to propose that words could be classified as clusters of sounds that occur together by suggesting that linguists could count the number of phonetic sounds that could follow a particular string in any language in order to determine the likelihood that the string occurred in or at the end of a word. Hayes and Clark (1970) suggested that learners use a clustering mechanism in which correlations between sounds within speech streams are calculated. They tested this explanation for segmentation by having participants listen to 45 min of a continuous and monotonous speech stream consisting of four nonsense words. It was found that participants could learn to distinguish between the nonsense words and completely novel strings of syllables with only the distribution of *phoneme*s as a cue to segmentation, although the effect was somewhat weak and the testing procedure left open the question of whether participants were identifying nonsense words based on exposure to the stimulus or the frequent repetition of them during the testing phase.

In one of the first studies directly addressing learners' use of transitional probabilities in language learning, Saffran et al. (1996) tested the ability of college students to segment six tri-syllabic nonsense words in the absence of all cues save transitional probabilities. The researchers created their nonsense words using four consonants (p, t, d, b) and three vowels (a, i, u) arranged into 12 syllables. The nonsense words were then pseudo-randomly *concatenated*, with the condition that no nonsense word could be repeated twice in a row. Each nonsense word was repeated 300 times, resulting in a speech stream of 21 min duration. Transitional probabilities between syllables within nonsense words ranged from .31 to 1, while those at word boundaries ranged from .1 to .2. Results showed that participants were able to distinguish exposure nonsense words from nonwords (completely novel strings of syllables) 76% of the time and part-words (reorderings of syllables from the nonsense words) 65% of the time, scores that are significantly better than would be expected by chance. Further, by grouping nonsense words according to high (.75 to 1) and low (.37 to .5) transitional probabilities, the authors found that nonsense words with high (79%) transitional probabilities were learned better than those with low (72%) transitional probabilities. Using identical techniques, but substituting tones for syllables, Saffran, Johnson, Aslin, and Newport (1999) showed that adults can also learn tonal "words" based on transitional probabilities.

In fact, transitional probabilities are such effective cues to word segmentation that Toro et al. (2005) and Mirman et al. (2008) reported that participants learned four nonsense words with perfect transitional probabilities (i.e., 1) within words and word boundary transitional probabilities of .33 at rates better than chance after only 7 min of exposure, a third of the time than in studies using imperfect transitional probabilities (i.e., less than 1). Schön et al. (2008) were able to replicate previous findings of participant learning after 21 min of exposure, but found that imperfect transitional probabilities are an insufficient cue for learners to segment speech streams containing six nonsense words after 7 min of exposure.

Computational Models

Computer models further support theories of transitional probability learning in language acquisition, although there have been some disagreements (Olds Batchelder, 2002; Perruchet & Peeremen, 2003; Perruchet & Vinter, 1998). For example Perruchet and Vinter (1998) suggested that there is a much simpler explanation for the previous findings than the conclusion that learners engage in statistical calculations. The researchers used the computer program PARSER to simulate word segmentation. PARSER was able to segment nonsense words equally well as participants in Saffran et al.'s 1996 study, but relying on basic principles of memory (repetition effects, decay and interference) and associative learning (unitization of elements processed in the same attentional focus) rather than computation of transitional probabilities. Based on this, Perruchet and Vinter (1998) argued that correct word segmentation is a direct result of the organization of the cognitive system. Individuals naturally chunk auditory stimuli based on spatial and temporal proximity into units of attentional focus that are then translated to mental representational units. The fate of each representational unit is decided by whether or not it is repeated; if a representational unit is a word, or a syllable within a word, that corresponds to a real life object or concept, it will be repeated. Repeated representational units are strengthened while those not repeated decay. A study by Perruchet and Peereman (2003) found that the PARSER model better predicted performance of human participants than another computational model based on automatic computation of statistical regularities, suggesting that sensitivity to statistical regularities occurs as a by-product of attention to input rather than being the primary analytic mechanism. Other researchers consider calculation of statistical properties as the primary mechanism of language acquisition. For example, Aslin et al. (1998) argued against a frequency of presentation explanation for their results with 8-month-old infants because it would involve the assumption that infant listening preferences varied randomly across studies. Thompson and Newport (2007) controlled for frequency of presentation with adults statistically and found support for the notion that calculation of transitional probabilities is distinct from mere frequency. Transitional probabilities have also been shown to be independent of other cues in the auditory stream, such as *phonotactic constraints* (Brent & Cartwright, 1996).

Segmentation and Meaning

Regardless of whether attentional processing leads to sensitivity to statistical properties of language or if learners start with calculations of transitional probabilities, they are useful in segmenting speech streams, and the segments extracted can be mapped to meaning. Mirman et al. (2008) paired transitional probability learning of a nonsense language with artificial lexicon training to investigate if segments identified using transitional probabilities could be used as meaningful object labels by adult participants. Results indicated that adults performed more poorly when asked to map labels to novel meanings if the transitional probabilities in those labels violated previous statistical learning. Graf Estes et al. (2007) found that transitional probabilities consistent with previous statistical learning facilitated mapping of labels to novel meanings for infants. These findings support the claim that statistical learning plays a role in actual language learning. Finally, Saffran (2001b) found that 8-month-old infants could learn word-like units using statistical learning, and that nonsense words were perceived as separate from the English contexts in which they were presented.

Phonotactic Constraints

Other cues to word segmentation are related to transitional probabilities; for example, because transitional probabilities are constrained by phonotactics, phonemes can also provide cues for segmentation (Hockema, 2006). According to Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk (1993), infants learn a great deal about the phonotactic constraints of their native language between 6 and 9 months of life. In their study, American and Dutch infants recognized words phonetically allowed by their native languages at 6 months while at 9 months infants showed preference for words phonotactically allowed by their native languages. Having learned about the allowable sound combinations in their native language, very young language learners are more apt to concentrate on syllable-level segmentation when attempting to understand fluent speech (Jusczyk, 1993). According to Jusczyk (1993), the level of discrimination may change as the individual's lexical corpus grows to include more confusable words that require the perception of *allophonic* differences to distinguish them, so that by adulthood individuals are able to derive phonemic representations of language input. This may be due, at least in part, to acquiring the necessary knowledge structures for reading. Jusczyk (1993) does argue, however, that phonemic discrimination may only be used in online language processing as a secondary probe to lexical memory.

Confirming that phonemic differences vary in importance depending on the developmental level of the learner, Cutler and Mehler (1993) argue that vowels are essentially more perceptually salient for infants by virtue of higher placement on the *sonority hierarchy* (see Table 1), as well as their typically greater duration and periodicity. The latter two characteristics of the vowel are due to the fact that vowels form the nuclei of syllables, whereas consonants are used in the *onset* or *coda* positions. These characteristics of vowels are more salient for young learners because they cohere to the rhythm of language, to which infants and children are biased to attend (Cutler & Mehler, 1993). Polka and Werker (1994) found that language-specific sound preferences developed at an earlier age for vowels than for consonants. As language development progresses, the learner develops and uses language-specific consonant discrimination (Werker & Tees, 1984). In addition, consonants may become more important than vowels for the older learner as demonstrated by Bonatti et al. (2005) in a series of experiments with native French speakers. Participants in these studies were unable to distinguish words from part-words when transitional probabilities were associated with vowels, but could do so when transitional probabilities were associated with consonants. Mehler, Peña, Nespor, and Bonatti (2006) found that consonants are more compatible with word learning based on transitional probabilities for adults and Cutler and Otake (2002) found that consonants exerted greater constraints on lexical identity than did vowels. Toro, Nespor, Mehler, and Bonatti (2008) clarify this relationship, by showing that adults use vowels to extract generalizations of structure while using consonants for word extraction.

Table 1

Sonority Hierarchy

Sound Type	Letter Type	Examples
Obstruent	plosives	b, d, g, p, t
	fricatives	s, f, v, z, h
Sonorant	nasals	n, m
	liquids	l,r
	high vowels	u, i
	non-high vowels	a, e

Note: table goes from lowest to highest in sonority

Onnis, Monaghan, Richmond, and Chater (2005) illustrated the importance of *phonology* in language acquisition with three studies conducted to resolve contradictory findings between Newport and Aslin (2004) and Peña, Bonatti, Nespor, and Mehler (2002). The Peña et al. study found that participants could learn on the basis of *non-adjacent dependencies* (e.g., the middle syllable of a tri-syllable nonsense word was random) while Newport and Aslin (2004) found that participants could not. Onnis et al. (2005) found that removing phonotactic similarity for nonadjacent syllables eliminated learning, resolving the contradiction by illustrating the phonological confound in the Peña et al. study: all first and third syllables began with *plosive* consonants (i.e., a consonant sound produced by stopping the airflow in the vocal tract), while second syllables began with *continuants* (i.e., a sound produced with an incomplete closure of the vocal tract). Supporting the phonological confound theory, Kessler and Treiman (1997) found that consonants are not equivalent in onset positions. Absolute phonotactic constraints guide categorization of a string as either legal or illegal, whereas probabilistic phonotactic constraints define the likelihood of a sound falling in a particular position in a word. For example, consonants lowest on the sonority hierarchy (plosives) are preferred as word onsets while those higher are less favored (Onnis et al., 2005) and more sonorous consonants are more cohesive with preceding vowels (Content, Kearns, and Frauenfelder, 2001). Similar effects were found by Creel, Newport, and Aslin (2004) who also looked at nonadjacent dependencies, but with tones, and found that similarity of tones provided the needed structure to learn coherent sequences despite temporal distance.

Although not directly addressed in the article, the two studies done by Saffran et al. (1996) also highlight the influence consonant type can have on learning by transitional probability. In their first experiment, the researchers used a nonsense vocabulary built on four consonants of low sonority: p, t, b, and d, while the second experiment's language was built on two consonants of low sonority and two of high: p, t, m, and n. In both experiments participants in the control conditions were able to learn with only transitional probabilities as cues to segmentation, but the difference in performance is telling. Participants in the first experiment distinguished words from the nonsense stream from nonwords 76% of the time, while those in the second experiment could only do so 65% of the time.

A large portion of the developmental differences seen in language acquisition at all levels can be attributed to previous experience with language, and so it is with phonotactics. Vitevitch and Luce (1998) found that probabilistic phonotactic learning was affected by *lexical competition*, such that non-words were more easily perceived because they did not have prior lexical associations, whereas actual words with high probabilities and high competition were harder to identify. In other words, phonotactic effects on learning are stronger when the learner is presented with a novel language than when one is learning new words in a native language because the new words are not initiating probes to lexical memory.

Prosody

Transitional probability learning may lead to knowledge regarding another property of speech, prosody (Cutler & Otake, 2002; Swingley, 2004; Thiessen & Saffran, 2003), although Jusczyk (1999) argued that the directionality was reversed, particularly for English learning. Prosody refers to the intonation, *pitch*, loudness, rate, and rhythm of speech. Prosodic cues are important for deciphering overall meaning of speech and can be used on a more fundamental level as cues to segmentation. Although mapping of prosody to word boundaries is often not consistent, Shukla, Nespor, and Mehler (2007) found that it can serve to filter out lexical candidates and infants can use prosody to segment clauses (Seidl & Cristià, 2008). Seidl and Cristià (2008) also found that infants' sensitivity to language specific prosodic cues rather than overall prosody develops between 4 and 6 months of age (Seidl & Cristià, 2008). Therefore, providing far more consistent prosodic cues to infants can facilitate language learning, which is seen as the purpose for infant-directed speech or motherese. Prosodic cues in motherese are enhanced by more consistent pauses, simpler intonation patterns, final segment lengthening, and fluctuation of fundamental frequency contours which overall produce a more informative *gestalten* for infants learning speech (Kemler Nelson et al., 1989).

According to McDonald (1997) both prosody and phonology are useful for determining the structure of language. Function and *content words*, for example, have different prosodic and phonotactic qualities that learners can exploit to properly categorize words segmented from speech (McDonald, 1997). Prosody can also be used for determining structure within words. Many words in English, for example, follow a trochee stress pattern (i.e., a stressed syllable followed by an unstressed one) that can help learners identify words. Echols, Crowhurst, and Childers (1997) found that 9-month-old infants preferred 2-syllable words where the trochee structure was preserved, but that 7-month-old infants did not. These results are confirmed by findings that 7-month-old infants attend to statistical cues while 9-month-old infants attend to stress cues when attempting segmentation (Thiessen & Saffran, 2003). Also in line with these findings, Houston, Santelmann, and Jusczyk (2004) found that 7.5-month-old infants could only segment tri-syllabic words from a speech stream when the first syllable received *primary stress*. In fact, Mattys, Jusczyk, Luce, and Morgan (1999) found that prosody can overrule phonology if the two conflict, at least for 9-month-old infants.

Saffran et al.'s (1996) second experiment with adults inserted a consistent prosodic cue, vowel lengthening, into speech streams containing six nonsense tri-syllabic words. Results indicated that lengthening of the final syllable facilitated segmentation, while lengthening of the first syllable impaired it. The authors indicate this may be due to exposure to English, as many English words have lengthened final syllables. Another explanation is that transitional probabilities are computed backwards, making word endings more salient. For example, Kempe, Brooks, Gillis, and Samson (2007) found that *diminutive endings* are a cue for word segmentation.

Structural Relationship Between Music and Language

Given that prosody involves the rhythm and intonation of language, and that the generalized mechanisms of statistical learning have been shown to be applicable to both tonal and language learning (Creel, Newport, & Aslin, 2004; Saffran et al., 1999), it is perhaps not surprising that some wonder at the nature of the relationship between language and music.

Patel (2006) discussed the relationship between musical and speech rhythms in terms of evolution of processes. He suggested that music is an innovation of the more "useful" cognitive process of language and that "grouping in music may well be an offshoot of prosodic grouping abilities" (p. 99). In support of a relationship between music and language, Patel and Daniele (2002) found that the prosody of a particular language can affect the structure of the instrumental music produced by the culture that speaks it.

Physiological studies also provide some support for a relationship between music and language. Traditionally, it has been noted that music is processed in the right hemisphere while language is processed in the left (Zatorre, Evans, Meyer, & Gjedde, 1992; reviewed in Peretz & Zatorre, 2005); however, recent research has shown that there is an overlap of activation in relevant brain regions (Brown, Martinez, & Parsons, 2006; Schön, Leigh Gordon, & Besson, 2005) involved in processing the structure of sounds but not the meaning (Besson & Schön, 2000; Peretz, Radeau, & Arguin, 2004), although "musical *semantics*" remains an ill-defined concept. Besson and Schön (2000) reviewed results of several studies of language and music processing that indicate that several important language areas are also involved in music; such areas include the primary auditory regions (BA 41/42) as well as secondary auditory regions (BA 22) and the supramarginal gyrus (BA 40). Overlap of activation for the basic processing and production of auditory stimuli is predictable, insofar as certain qualities are shared across specific types of sounds. What needs to be further clarified is what qualities are distinct to each type of auditory input. For instance, it appears that the processing of pitch and beat in music is distinct from language processing (reviewed in Peretz & Zatorre, 2005); Zatorre et al. (1992) found that processing of phonetic structure was related to activity in left hemispheric Broca's area, while pitch processing was related to activity in the right prefrontal cortex. Further clarification of the shared and distinct qualities of language and music, as well as the brain regions involved in their processing and production, is needed to make any firm conclusions about the relationship between the two types of input, but we can already see that there is a relationship to be explored.

Music and Word Segmentation

Building on the previous evidence of a relationship between music and language, it has been argued that music may contribute to language learning by enhancing phonological discrimination, increasing attention and arousal, and by optimizing learning mechanisms through redundant structural properties (Schön et al., 2008). In three experiments, Schön et al. (2008) exposed a total of 78 native French-speaking participants to experimentally manipulated speech streams designed to explore the possible facilitation effects of music on new language acquisition. While previous research had shown that participants could learn nonsense words based on transitional probabilities alone after 21 min of exposure (Bonatti et al., 2005; Mirman et al., 2008; Saffran et al., 1996; Saffran et al., 1997; Toro et al., 2005), the authors conducted their first experiment to determine if decreasing the exposure time by two-thirds would decrease learning. Six nonsense words were created by substituting French consonants and vowels into the six nonsense words used by Saffran et al. (1996). These nonsense words were then concatenated into a constant stream with no *segmentation cues* aside from the transitional probabilities, which ranged from .31 to 1 within words and from .1 to .2 between words. The six nonsense words were each repeated 108 times in the speech stream, so that the stream was approximately 7 min in length. The authors found that after this amount of exposure to a monotonic speech stream with only transitional probabilities to cue segmentation, the 26 participants performed no better than chance on nonsense word recognition.

In their second experiment, Schön et al. (2008) used the same speech stream they had created for Experiment 1. In this case, however, the authors assigned a particular tone to each syllable in the speech stream on which it was consistently "sung" by the synthesizer throughout the exposure period. Care was taken to have no *pitch contour* changes within nonsense words, while the chance was approximately 50/50 of a contour change between nonsense words; thus, contour could not be used as a segmentation cue. Under these conditions, 26 participants were able to discriminate the six nonsense words from part-words 64% of the time after only 7 min of exposure, leading the authors to conclude that the addition of musical information does indeed assist participants in discerning word boundaries.

The final experiment of Schön et al. (2008) was designed to address why facilitation occurred. The authors wanted to find out if the gains in learning were due to an overall increase in arousal and attention, boundary enhancement due to the gestalt properties of pitch, or enhancement of global transitional probabilities. Twenty-six participants heard a stream of "sung" syllables of the same structure found in Experiment 2, except the musical and linguistic boundaries did not overlap because the musical structure was moved one step to the right. This resulted in the second and third syllable of each nonsense word being "sung" on consistent pitches, while the first syllable was randomly varied amongst six pitches. Participants were able to distinguish nonsense words from part-words significantly more often than expected by chance (56%), yet less than was found in the previous experiment. These results suggest that superposed transitional probabilities are important to tonal facilitation effects on learning, but that boundary enhancement and arousal effects also contribute. The authors also briefly suggest in their discussion that music is akin to prosody, but this is not directly tested.

Overview of the Current Study

The current research was designed to serve several purposes. First, replication of the results of Schön et al.'s (2008) first experiment was intended to illustrate that transitional probabilities alone are not sufficient for English-speaking participants to learn the nonsense words with an exposure period of only 7 min. This group of participants then served as the comparison group for three experimental conditions: prosody-enhanced, tonally-enhanced, and tonal-word. These conditions were used to assess tonal enhancement of language acquisition as well as to compare tonal enhancement to prosodic enhancement to determine if musical information may indeed be considered akin to prosody or if there were distinct effects of the two types of cues.

The six trisyllabic nonsense words used in this study were taken from Saffran et al.'s (1996) second experiment, both because the difference in performance between experiment 1 and experiment 2 in that paper suggest a possible confound of consonant type, and also because these words were used, with slight adaptation to French sounds, for Schön et al.'s (2008) research. The six trisyllabic words are: mupana, mamupu, nutama, patumi, pinamu, and tutimu. Thus it can be seen that three words begin with consonants lower in sonority, while three begin with consonants higher in sonority. The use of these words was also meant to preserve the transitional probabilities of previous studies, which ranged from .31 to 1 within words and from .1 to .2 between words.

After exposure, participants performed a forced-choice task, in which they were asked to choose between a trisyllabic part-word created by rearranging syllables from the nonsense words, and actual nonsense words from the exposure stream. In addition, half the participants in each condition heard test stimuli presented in monotone speech, consistent with previous literature, but the other half heard test stimuli presented as they were heard during the exposure period. This was intended not only to obtain purer estimates of learning, but also to provide some information about generalization of newly learned sounds. Jusczyk (1997) discusses the development of infants' ability to generalize amongst various voices speaking the same words, suggesting that the initial lack of this ability is due to limited memory representations. However, because the nonsense language may activate tactics for segmentation used earlier in development due to the lack of lexical competition (Vitevitch & Luce, 1998), it is important to assess if participants can identify nonsense words more easily if they are presented for testing as they were in familiarization.

The first experimental condition was tonally-enhanced just as in the Schön et al. (2008) article in order for comparison with the control group, to replicate the tonal facilitation effects with English speaking participants. The next experimental condition contained prosody enhancement. Specifically, the vowel of the word-final syllable was lengthened by 100 ms. The third experimental condition consisted of a tonal-word structure. Thus the current research had several hypotheses.

Hypotheses

Hypothesis 1: Participants in the control condition will perform no better than chance when tested with part-words after 7 min of exposure to the speech stream.

Hypothesis 2: Participants in the tonally-enhanced condition will perform better than chance when tested with part-words after 7 min of exposure to the speech stream.

Hypothesis 3a: Participants in the prosody-enhanced condition will perform better than chance when tested with part-words after 7 min of exposure to the speech stream.

Hypothesis 3b: Participants in the prosody-enhanced condition will perform better than participants in the tonally-enhanced condition.

Hypothesis 4a: Participants in the tonal-word structure condition will perform better than chance when tested with part-words after 7 min of exposure to the speech stream.

Hypothesis 4b: Participants in the tonal-word structure condition will perform better than those in the tonally-enhanced condition, and equal to those in the prosody-enhanced condition.

Hypothesis 5: In the three experimental conditions, participants will perform better when test trials are presented in the same fashion as the exposure period than when test stimuli are presented in a different fashion.

METHODS

Participants

Participants were 86 undergraduate male and female students at the University of Wisconsin Oshkosh who were enrolled in psychology classes and participated for class credit. Of these, three participants were dropped from final data analyses; one was dropped because she did not appear to understand either the verbal instructions or the written questionnaires, and two because they were in the prosody-enhanced condition, designed for native English speakers, but English was not their native language. This left 19 participants in the prosody-enhanced condition, while there were 22 in the control, and 21 each in the tonal-word and tonally-enhanced conditions. Only four non-native English speakers were included in final analyses; three of these were native Hmong speakers while the fourth was a native French/Arabic speaker. These participants were in conditions other than the prosody condition and because those streams were not designed to be English-specific, their native languages should not have significantly affected their learning. This was confirmed by the finding that removing these participants from the analyses did not change results. Overall, 36 participants had some level of experience with a language other than their native one, while 47 did not. The mean duration of other language experience was 5.30 years (SD = 4.14) with a maximum of 18. Participants ranged in age from 17 to 43, with a mean of 20.3 (SD = 3.7). There were 46 female and 37 male participants. Fifty-four participants (65%) had some musical training; of these, 36 had instrumental training alone, 3 had vocal training alone, and 14 had both kinds of musical training. The mean duration of musical training for these participants was 5.88 years (*SD* = 3.36) with a maximum value of 15. All participants reported normal hearing or corrected to normal hearing. (See Appendix D).

Materials

Speech streams were synthesized in an American male voice using the Mbrola speech synthesizer (http://tcts.fpms.ac.be/synthesis/mbrola.html). Four consonants (m, n, p, t) and three vowels (a, i, u) were used to create 11 nonsense syllables that were then be grouped to create six tri-syllabic words (mupana, mamupu, nutama, patumi, pinamu, tutimu). A speech stream was created for each of four conditions. Each stream consisted of 108 presentations of each nonsense word in pseudo-random order, with the stipulation that no word be presented twice in a row. For the control, tonal-word, and tonally-enhanced conditions, the speech stream was 6 minutes 29 seconds long, while the prosody-enhanced stream was slightly longer, at 7 minutes 34 seconds. Transitional probabilities were first assumed to be as reported in the previous literature (.1 to .2 between words and .31 to 1 within words), given that the nonsense words and manner of concatenation were the same, but calculation revealed that two between-word transitional probabilities were well above the desired range, at .42 (mu nu) and .66 (mu pa); see the Discussion section for possible implications of these high transitional probabilities. Within-word transitional probabilities were consistent with those reported in previous studies, ranging from .31 to 1.

For the control condition, the six nonsense words were concatenated into a monotonic stream, such that the only cue available for word segmentation was the transitional probabilities within and between words. For the first experimental condition, the prosody-enhanced condition, the third-syllable vowel of each of the six words was lengthened by 100 ms, but the speech stream was otherwise identical to the control condition. In the speech stream for the second experimental condition, the tonally-enhanced condition, each of the 11 syllables was assigned a tone (C5, D5, F5, G5, A5, B5, C6, Db6, D6, E6, and F6)[2] on which it was consistently "sung". Each nonsense word was thus 'sung' on a specific melodic contour each time it was presented. Contour did not change within words, but this consistency was not a cue to segmentation because there were roughly equal numbers of continuous and discontinuous contours at word transitions. In other words, a word with a rising contour was roughly equally likely to be followed by either a higher or lower pitch. In addition, the mean pitch interval within words was roughly equivalent to that between words. The fourth speech stream, the tonal-word condition, also used tones in conjunction with the nonsense words but rather than give each syllable a distinct tone, there was a set word structure. Specifically, the first and second syllables of every word token were "sung" on the same tone (A4), while the last syllable of every word token was "sung" on a different tone (C4).

---

[2] The letter refers to the note, the number to the *octave*, and "b" indicates a flat note. So Db6 is D flat in the 6th octave of the 88-key piano keyboard.

Procedure

In individual sessions participants were randomly assigned to one of the four conditions and then asked to listen to one of the speech streams presented via headphones; participants were allowed to control the volume. The researcher instructed the participants to attend to the stimulus as they would be asked questions about it later, but did not give additional information about the stimulus (see Appendix C). Once the exposure period concluded, participants were asked to complete a forced-choice test where test items were presented auditorily and participants recorded answers on a form (Appendix E). Participants were alone in the testing room for both the exposure period and the testing phase and were in control of when the sound files began to play. Sound files were in a .wav format and played in Microsoft Media Player on a Compaq Evo N800v notebook laptop.

The forced-choice test items were of 36 pairs of nonsense words and part-word foils. The part-word foils consisted of two syllables from one of the nonsense words combined with an additional syllable from a different nonsense word as follows: mupatu, mamunu, namupi, patuna, tamapu, and timumi. The test pairs were presented 300 ms apart and there was a 3 s gap between presentations of pairs. For each pair, participants were asked to indicate which syllable string sounded more like a word from the exposure speech stream. Participants wrote “1” on the form to indicate the first word, and “2” to indicate the second word. For half the participants in each experimental condition, syllable strings in the testing phase were presented in a spoken monotone; the other half of participants heard the test syllable strings presented in the same manner heard in the exposure speech stream. For example, half the participants in the prosody condition were presented with test stimuli with lengthened final vowels while the other half heard test stimuli presented in a monotone with equal vowel lengths.

After completing the forced-choice test, participants were asked to complete the participant information questionnaire (Appendix D). Finally, the researcher explained the purpose and design of the study to participants, asked if they had any questions, and finally thanked and dismissed them.

RESULTS

Grouped Nonsense Words

The mean percentage correct expected by chance for a forced-choice test of this nature is 50% (18 out of 36). Percent correct was calculated for all four conditions and then tested against this predicted population mean to determine if participants' performance was better than chance. Participants in the control condition scored an average of 18.05 (SD = 3.05), or 50% correct. Participants in the tonal-word and tonally-enhanced conditions scored an average of 51% (M = 18.38, SD =3.97) and 49% (M = 17.67, SD = 3.51) respectively. In the final condition, the prosody-enhanced condition, participants scored an average of 63% (M = 22.53, SD = 4.71) correct. The only condition in which participants performed significantly better than chance was the prosody-enhanced condition, $t(18) = 4.19$, $p < .001$. Performance in the other three conditions was not significantly different from chance, all ts < 1, and all ps > .05.

An alpha level of .05 was used for all statistical tests. Distribution of scores was normal (Kolmogorov-Smirnov = .07, $p > .20$) and had homogeneous variance (Levene's = 7.82, $p = .15$). A one-factor (condition) between-subjects analysis of variance (ANOVA) revealed that there was a significant difference between conditions, $F(3, 79) = 6.87$, $p < .001$, $\eta^2 = .21$. Planned comparisons showed that performance in the prosody-enhanced condition was significantly better than in the control ($t(39) = -3.67$, $p < .001$, one-tailed), tonal-word ($t(38) = 3.02$, $p = .002$, one-tailed) and tonally-enhanced ($t(38)$ =

3.72, $p < .001$, one-tailed) conditions. No other planned comparisons showed significant differences between groups (all *t*s < 1, $p > .05$).

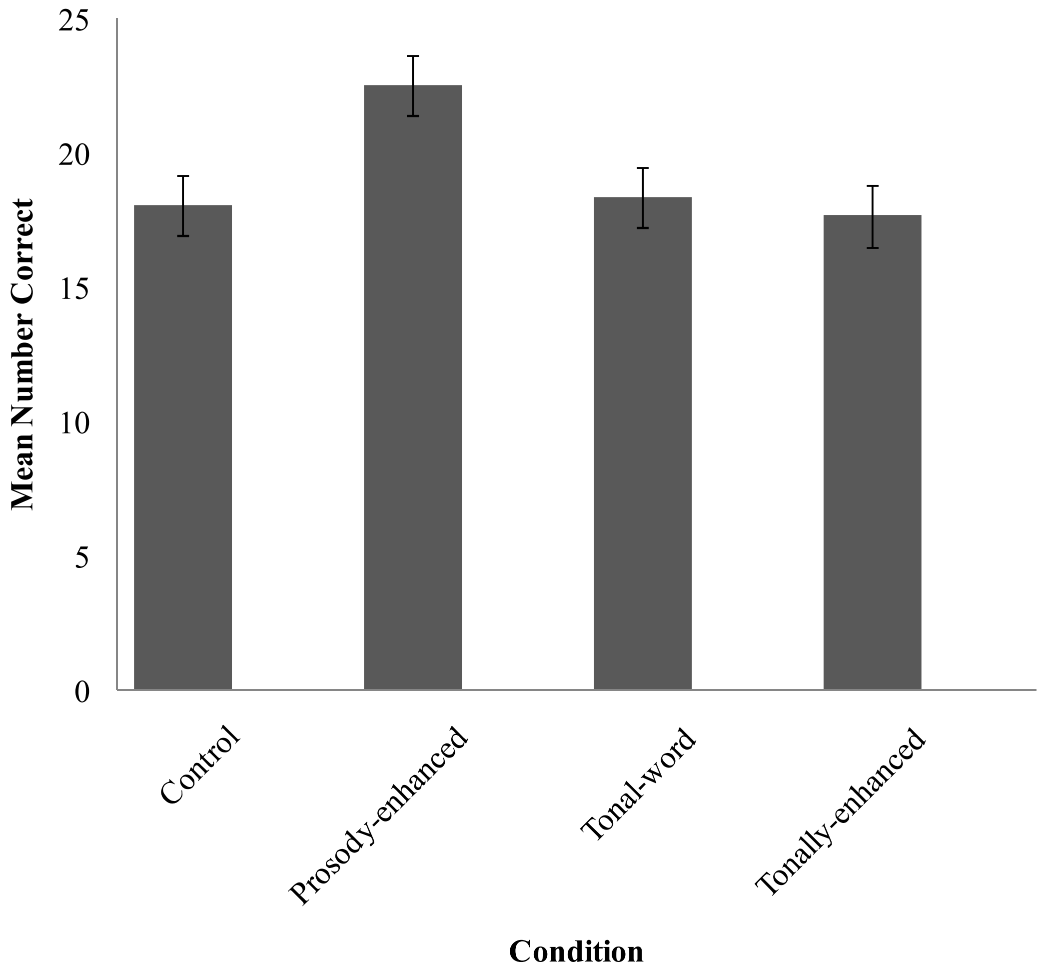

Figure 1. Mean number correct by condition.

In addition to the planned comparisons combining scores on all six nonsense words, the nonsense words were divided into two phonotactic subsets that were then compared for mean performance in all conditions, as well as in each condition individually. Across all conditions, participants scored an average of 9.12 (*SD* = 2.56) correct on the first subset (mupana, mamupu, nutama) and an average of 9.94 (*SD* = 3.01) on the second subset (patumi, pinamu, tutimu). A paired-samples *t*-test showed that this difference was significant ($t(82) = -2.04$, $p = .04$, two-tailed). Performance on the first subset was not significantly different than chance (9 out of 18) ($t(82) = .43$ $p = .67$, two-tailed), but performance on the second subset was ($t(82) = 2.85$, $p = .006$, two-tailed). Specifically, comparisons show that the prosody-enhanced condition performed better than the control ($t\ (39) = 3.94$, $p < .001$), tonal-word ($t\ (38) = 2.98$, $p = .005$), and tonally-enhanced conditions ($t\ (38) = 4.62$, $p < .001$) on subset 2, but performance was equal between conditions on subset 1 as follows: prosody by control ($t(39) = 1.41$, $p = .17$); prosody by tonal-word ($t\ (38) = 1.55$, $p = .13$); prosody by tonally-enhanced ($t(38) = 1.59$, $p = .12$).

There was a significant difference between performance on the first ($M = 10.10$, $SD = 2.66$) and second ($M = 8.96$, $SD = 2.21$) halves of the forced-choice test ($t(82) = 4.18$, $p < .001$). This is perhaps explained by the fact that several participants reported that after a few repetitions of the test material they began to get confused. Follow-up *t*-tests show that the effects of the prosody-enhanced condition were consistent, with participants performing significantly better than the control ($t(39) = -3.16$, $p = .003$), tonal-word ($t(38) = 2.38$, $p = .02$), and tonally-enhanced conditions ($t(38) = 3.56$, $p = .001$) on the first half as well as on the second half (prosody vs. control condition $t(39) = 3.07$, $p = .004$; prosody vs. tonal-word $t(38) = 2.99$, $p < .005$; prosody vs. tonally-enhanced $t(38) = 2.80$, $p = .008$).

Consistent with Saffran et al. (1996), nonsense words were also grouped into two subsets with the lowest mean TP[3] (transitional probability) (.63, .58, .38) and the highest mean TP (.75, .75, .65) subsets for comparison. Across conditions, it was found that the low mean TP subset (*M* = 9.94, *SD* = 3.03) scored better than the high mean TP subset (*M* = 9.12, *SD* = 2.39), with $t(82) = 2.15$ and $p = .04$. This result is likely due to the unique combination of the three words and the fact that there was not as large a spread of mean TPs as is usually reported in other studies. A one-factor (condition) between-subjects ANOVA revealed that there were no significant group differences for the low mean TP subset ($F(3, 79) = 2.65$, $p = .06$), but there were significant group differences for the high mean TP subset ($F(3, 79) = 7.33$, $p < .001$). Planned comparisons for the high mean TP subset showed that the prosody-enhanced condition performed significantly better than the control ($t(39) = 4.48$, $p < .001$), tonal-word ($t(38) = 3.80$, $p = .001$), and tonally-enhanced ($t(38) = 3.26$, $p = .002$) conditions.

Finally, performance across conditions was compared between types of presentation of the test-stimuli (the same as or different from the manner of exposure stream presentation). The mean for "same" test presentation was 18.80 (*SD* = 3.62) while that for "different" presentation was 19.45 (*SD* = 5.02); the difference between test presentations was not significant ($t(81) = .65$, $p = .26$, one-tailed).

---

[3] Mean transitional probability is the mean of the two TPs between the syllables within the word (first and second; second and third).

Individual Nonsense Words

Table 2 displays the mean number correct and standard deviation for each nonsense word across experimental conditions, as well as *t* scores representing a comparison between the achieved mean and the mean expected by chance. Chance performance for each word was again equal to 50%, or in the case of each word individually, three out of six.

Table 2

Across Condition Performance

Word	*M*	*SD*	*t*
mupana	3.53	1.28	3.79*
mamupu	3.27	1.52	1.59
nutama	2.33	1.43	-4.19**
patumi	3.53	1.53	3.12*
pinamu	3.16	1.60	-0.02
tutimu	3.25	1.54	1.42

* p < .01, ** p < .001

As can be seen in Table 2, when means were pooled across conditions participants performed better than chance on only two nonsense words: mupana and 4 patumi. Participants actually performed significantly worse than chance on the nonsense word nutama. Planned comparisons revealed that performance on nutama was significantly worse than performance on all other nonsense words (all *t*s > 3.77, all *p*s < .001). The only other significant difference found in the planned comparisons of means across conditions was that between mupana and pinamu with a *t*-score of 1.97 (*p* = .05).

When each nonsense word is compared across conditions, the results are somewhat different. Table 3 shows the means, standard deviations, and *t* scores comparing performance to that expected by chance for each nonsense word within each condition.

Table 3

Performance on Individual Nonsense Words

Word	Condition	*M*	*SD*	*t*
mupana	1	3.77	1.45	2.48*
	2	3.47	1.26	1.62
	3	3.76	1.14	3.04**
	4	3.10	1.22	0.37
mamupu	1	3.05	1.33	0.18
	2	3.79	1.44	2.39*
	3	2.67	1.43	-1.06
	4	3.62	1.72	1.63
nutama	1	2.14	1.36	-2.97**
	2	2.84	1.64	-0.42
	3	2.38	1.28	-2.21*
	4	2.00	1.41	-3.23**
patumi	1	3.00	1.69	0.00
	2	4.47	1.12	5.65***
	3	3.52	1.63	1.44
	4	3.24	1.26	0.86
pinamu	1	2.86	1.64	-0.40
	2	3.95	1.78	2.32*
	3	3.19	1.47	0.59
	4	2.71	1.31	-1.00
tutimu	1	3.23	1.31	0.82
	2	4.00	1.49	2.94**
	3	2.86	1.71	0.38
	4	3.00	1.48	0.00

Note: Condition 1 = control, 2 = prosody, 3 = tonal-word, 4 = tonally-enhanced
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 4 shows the results of one-factor ANOVAs for each word. Only word 4 (patumi) showed significant group differences. Follow-up *t*-tests showed that participants in the prosody-enhanced condition performed significantly better than those in the control ($t(39) = 3.23$, $p = .003$), tonal-word ($t(38) = 2.12$, $p = .04$), and tonally-enhanced ($t(38) = 3.26$, $p = .002$) conditions.

Table 4

ANOVA Results

Word	*F*	*p*
mupana	1.33	0.27
mamupu	2.49	0.07
nutama	1.34	0.27
patumi	3.91	0.01
pinamu	2.47	0.07
tutimu	2.25	0.09

Other Predictors

In addition to performance data, the Participant Information questionnaire (Appendix D) asked for several pieces of information that were considered potential predictors of performance. Years of both multiple language experience and musical training were tested for effects on performance. Analysis of covariance (ANCOVA) showed that neither years of language experience ($F(1, 74) = .02$, $p = .97$) nor years of musical training ($F(1, 77) = .54$, $p = .46$) contributed significantly to the variance attributable to condition.

DISCUSSION

In all, seven hypotheses were tested by this study. Of those, only three were supported. First, consistent with previous research, this study showed that adult learners cannot segment six tri-syllabic nonsense words from a continuous speech stream in under 7 minutes on the basis of transitional probabilities alone (hypothesis 1). Second, prosody-enhancement allowed learners to achieve overall performance scores that were better than chance on the forced-choice test (hypothesis 3a). Finally, participants in the prosody-enhanced condition performed significantly better than those in the tonally-enhanced condition (hypothesis 3b). Neither the tonal-word nor the tonally-enhanced conditions allowed participants to perform better than chance (hypotheses 2 and 4a) and both were inferior to performance in the prosody-enhanced condition (hypothesis 4b). Test presentation had no effect on performance, such that participants who were presented test stimuli in a manner congruent with the stimulus presented during the exposure period performed equally well as participants who were presented test stimuli in an incongruent manner (hypothesis 5).

These results also revealed that overall group differences could be attributed to only three of the nonsense words: mupana, nutama and patumi. Specifically, mupana and patumi were facilitated by the prosody-enhanced condition, but nutama was inhibited. This pattern was not expected and is difficult to account for when examining all words across conditions, but examining performance on different subsets as well as on individual words within conditions may shed light on this pattern of results.

Nonsense words were separated into two subsets based on the probabilistic phonotactics of their initial letters. The subset consisting of nonsense words that began with plosive consonants was learned significantly better than the subset consisting of nonsense words beginning with *nasal* consonants, although only one word (patumi) in the plosive subset was learned significantly better than chance when considered alone. However, performance in the prosody-enhanced condition (Condition 2 in Table 3) alone is further indicative of learning. Participants in the prosody-enhanced condition learned all three plosive-initial nonsense words as well as one of the nasal-initial nonsense words at better than chance levels. Perhaps more important, while all three of the other conditions showed impaired performance on nutama, performance in the prosody-enhanced condition was at chance level. Thus, prosody-enhancement improved performance on 5 out of 6 nonsense words, illustrating the power of this cue to word segmentation.

The control condition had two significant findings when individual words were compared between conditions; this is contrary to expectation, based on previous research that performance would be at chance levels. Mupana was learned better than chance, while nutama was actually learned worse than chance. To explain this, it is necessary to take into account other characteristics of the nonsense word. First, mupana was the first nonsense word in all the exposure streams, which could have been a significant cue used to learn the word. In addition, mupana contained the most often repeated syllable in the entire stream, “mu”, meaning that the mean transitional probability of this word, while not one of the highest, was also not low enough to impair learning. The between-word transitional probability for combinations ending in “mu” was also low enough not to impair learning. Finally, mupana contains the phonotactic similarity between the first and third syllables that Onnis et al. (2005) found necessary for adjacent probability learning. Nutama also had phonotactic similarity working for it, and its mean within-word transitional probability was even stronger than that for mupana. However, the between-word probability for combinations ending in “nu” was much higher than intended, at .42. This fact, combined with the bias against nasal consonants in onset positions, may account for the impaired performance on this word, not only in the control condition but also in the tonal-word and tonally-enhanced conditions. Performance in the control condition met with expectations for the other four nonsense words.

In the tonal-word condition (Condition 3 in Table 3) only mupana was learned better than chance and performance on nutama was impaired; both of these results are likely due to the reasons discussed above. The tonal-word condition thus provided no facilitation to word segmentation. The tonally-enhanced condition also provided no facilitation to learning any of the nonsense words, and may even have impaired learning of mupana, since even the prosody-enhanced condition showed somewhat better, although not significantly better, learning.

The current study failed to find any facilitory effects of tone on word segmentation. All results of the study indicate that tones did not provide any facilitation equal to or greater than facilitation provided by a regular speech cue. However, the regular speech cue did prove to have impressive facilitory effects. Consider that Saffran et al.’s (1996) second experiment, which also used this cue, provided 21 min of exposure to speech streams, yet in just a third of the learning time, participants in this study performed as well as those in the Saffran et al. study, showing facilitation for 5 of 6 nonsense words.

The tonal-word condition was designed to test the idea that any type of repeated pattern added to the speech stream could facilitate learning. It was hypothesized that this condition would produce results equal to that of the prosody-enhanced condition because of the similarity of the patterns: first and second syllables are pronounced the same, while the third is pronounced differently. Clearly the results do not support this idea, instead indicating that prosodic cues found in one's native language have a privileged status when it comes to processing unfamiliar language input.

The failure of the tonally-enhanced condition to facilitate learning is more puzzling. Saffran et al. (1999) showed that it was indeed possible for adults to learn tonal "words" based on transitional probabilities, just as they can learn nonsense words, and Schön et al. (2008) were able to use tones to facilitate learning. Schön et al. found reason to conclude that tones facilitated learning for three reasons: reinforcement of transitional probabilities, boundary enhancement, and arousal. Perhaps the failure to find an effect in this study can be explained in these terms.

First, while reinforcing the transitional probabilities of the nonsense words, the tones were also reinforcing the two unexpectedly high between-word transitional probabilities (.66 for 'mu pa' and .42 for 'mu nu'). It is understandable that this reinforcement of bad information could impair performance. The problem with this explanation, however, is that every syllable in the current study corresponds to a syllable in the Schön et al. (2008) study, and likewise with the actual nonsense words. Also, tones listed by Schön et al. were applied to the English syllables that directly corresponded to the French syllables. However, a potentially relevant difference is that the nonsense words of Schön et al. all began with consonants of low sonority and such consonants have been reported to be preferred word onsets. Boundary enhancement is another proposed function of tones added to syllables, such that Schön et al. found a lower rate of learning when the musical boundary was shifted so as to be distinct from the linguistic boundary of the nonsense words in their third experiment. However, given that no consistent structural changes in tone were cues to segmentation, and that the two high between-word probabilities were reinforced, it seems unlikely that the tonally-enhanced condition was capable of enhancing boundaries. The final explanation for tonal facilitation is that musical information increases arousal, according to Schön et al., but their series of studies kept arousal constant rather than measuring it directly. This study was likewise incapable of directly assessing arousal. One participant did make a point to state that he found the tonally-enhanced speech stream particularly annoying, but this is hardly a basis for drawing any conclusions for the entire condition.

Given the previously discussed problems, no concrete conclusions can be drawn as to why the tonal information was unsuccessful at facilitating word segmentation in this study, but there appear to be two main possibilities. First, there may have been vital differences between the stimuli in this study and those studies up on which it was based. Given that the transitional probability appears to be a fairly basic cue, susceptible to the influence of more complex cues such as phonology and prosody, and that transitional probabilities are likely to be closer in value between and within words in natural language (Saffran et al., 1996), a decision was made to try to eliminate phonological confounds while following all other reported stimuli production procedures. This decision resulted in the two high between-word transitional probabilities that were mentioned previously and likely contributed to the failure of the tonally-enhanced condition to facilitate learning. Secondly, there is the possibility that tones simply do not produce the same facilitation effects for English speakers that they do for French speakers, but this conclusion is premature and certainly deserves further exploration.

With respect to other potential influences on performance, no specific hypotheses were made, yet intuitively one might expect that previous experience with multiple languages would improve performance on processing of unfamiliar language input and that previous musical training would be beneficial when that input also included tonal information. Statistical tests did not confirm either of these intuitions. One reason for this might be that the questions on training were not specific enough. It is customary for children attending public school to receive opportunities for musical and language training, but there are large differences in the quality of music instruction, (Rauscher, 2005), and not all students choose to continue these pursuits. In the current study, for example, a participant may have responded that she had a year of musical training, yet this tells us nothing about when this training occurred. Given the body of research suggesting that the non-musical benefits of music instruction are evident only in people who began music instruction before age 6 or 7 (for review, see Rauscher, 2008), such information would perhaps prove vital to determining effects of music training on linguistic processing.

This study would be improved by ensuring that the transitional probabilities for all the nonsense words fall within the same range, while refraining from introducing confounds of phonology or regularity. The current results can speak only to the power of the language-specific prosodic cue, final vowel-lengthening, in word segmentation. The value of adding musical or tonal information to language input remains to be accurately measured for English speakers.

Future Research

Because most studies of the relationship between music and cognition point to the fact that it is musical training rather than mere exposure to music that contributes to observed cognitive differences, further research on the link between language and music should include more detailed examination of music training and aptitude. Also of interest would be studies looking at more complicated musical and linguistic structure in order to clarify areas of overlap and distinction of the two domains. Beyond the simple constructions of tones present in this study, more song-like structures might assist linguistic processing at other levels, such as syntax and grammar. It also may be the case that musical information may help learners overcome barriers to second language learning by enriching language input. Creating more realistic learning environments and examining the effects of musical enrichment might eventually be used to develop educational programs to assist in the acquisition of new languages for older learners by detailing what sorts of enrichment are effective at overcoming barriers to such learning. For example, perhaps musical information may help overcome learner biases for native language cues, such as that seen in the prosody-enhanced condition of the current study.

Another question for future research is what effect tonal/musical information could have on initial language learning, as opposed to second language learning. This study, while not explicitly about second language learning, used a sample of adults who have already acquired at least one language, but there are good reasons to believe that the mechanisms for learning a first vs. additional languages are similar but distinct, perhaps because of differences in general cognitive mechanisms such as processing capacity (Cochran, McDonald, & Parault, 1999). It may be that a linguistic environment enriched with musical information would prove advantageous for initial language learning.

Music-language studies such as this one are important because they suggest ways of bridging the current divide between the sciences and the humanities. Prominent minds on both sides of this divide are advocating for studies that bring these two frameworks of human knowledge together (e.g., Wilson, 1998; Becker, 2004; Edelman, 2006). The study of music-language relations is one area in which scientific and humanistic studies can meaningfully intertwine, and in which interactions across traditional boundaries can bear fruit in the form of new ideas and discoveries that neither side can accomplish alone. Although studies that unify scientific and humanistic knowledge are still uncommon, comparing music and language provides a powerful way to study the mechanisms that the mind uses to make sense out of sound. Such studies potentially have implications for both practical and theoretical issues surrounding human communicative development.

APPENDIX A

Glossary

Allophone	alternate pronunciation of a phoneme that is dependent upon the phoneme's position within a word
Coda:	the consonant sounds of a syllable that follows the nucleus
Concatenated:	put together in a string
Content Word:	word that conveys meaning, such as a noun or verb
Continuant:	consonant where air continues to be released after articulation
Diminutive Endings:	formation of a word used to convey a slight degree of the root meaning, smallness of the object or quality named, encapsulation, intimacy, or endearment
Fricative:	consonant sound made by forcing air through a narrow channel
High Vowel:	vowel produced with the tongue at the roof of the mouth
Imperfect TP:	less than 1 transitional probability (TP); occurs when one syllable may follow another in a word and across word boundaries
Lexical Competition:	simultaneous cognitive activation of several similar word candidates
Liquid:	trill, tap, or approximate consonant that is not classified as a semivowel (glide) because it does not correspond phonetically to a specific vowel
Nasal:	consonant produced with a lowered velum in the mouth, allowing air to escape freely through the nose
Non-adjacent dependencies:	dependencies between two segments separated by a random middle segment
Non-high Vowel:	vowel produced with the tongue far from the roof of the mouth
Obstruent:	consonant sound formed by *obstructing* airflow, causing increased air pressure in the vocal tract.
Octave:	the interval between one musical pitch and another with half or double its frequency
Onset:	consonant, consonant cluster, or null space at the beginning of a syllable

Perfect TP:	Transitional probability (TP) of 1; occurs when one syllable always follows another within a word and never across word boundaries
Phoneme:	smallest structural unit that distinguishes meaning, though without semantic content itself
Phonology:	language-specific systems and patterns of sound and gesture
Phonetics:	the physical sounds of human speech
Phonotactic constraints:	permissible combinations of phonemes
Pitch:	the property of a sound or musical tone measured by its perceived frequency
Pitch Contour:	a function or curve that tracks the perceived pitch of a sound over time
Plosive:	consonant produced by an explosive release of air from the mouth
Primary Stress:	strongest degree of stress placed on a syllable in a word
Prosody:	the patterns of stress and intonation in a language
Segmentation Cue:	characteristics of speech input used to identify boundaries
Semantic:	the study of meaning of communication
Sonorant:	a speech sound that is produced without turbulent airflow in the vocal tract
Sonority Hierarchy:	ranking of speech sounds by amplitude
Speech Segmentation:	the process of identifying the boundaries between words, syllables, or phonemes in spoken natural languages
Transitional Probability (TP):	the likelihood that one segment will follow another

APPENDIX B

Notes on Infant Research

Infant studies of language acquisition are different from the adult studies summarized above in several ways. First, infants are typically familiarized with fewer nonsense words (four vs. six) for shorter periods (2 to 3 min vs. 21 min) than adults to accommodate their shorter attention spans. Next, because of their inability to complete forced-choice test trials, the dependent measure in infant studies is looking time. For testing, the infant is seated on the lap of a caregiver in sound-attenuated booths, with researchers and cameras positioned outside the booths to observe the direction of the infant's gaze. Inside the booths, speakers and lights are positioned in front and to either side of the infant. Flashing lights are used to draw the infant's attention to a given direction, then the audio stimuli are played from the corresponding speaker. Trials end when the infant's looking time deteriorates to a predetermined level. Both researcher and caregiver are blinded to test stimuli with masking headphones. When the infant's looking times are longer for test stimuli not heard or heard less frequently in the habituation session, the infant is said to show a "novelty preference." If looking times are longer for test stimuli heard more frequently in the habituation session, it is termed a "familiarity preference." At different ages, infants display different listening preferences. For example, Thiessen and Saffran (2003) found that 7-month-olds showed a novelty preference, while 9-month-olds showed a familiarity preference.

APPENDIX C

Participant Instructions

Initial Instructions

“In this study, you will be asked to listen to computerized sounds for 7 min, afterwards you will answer questions about the sounds. There will also be a short questionnaire asking for information related to the study.”

Listening Phase Instructions

“For the next 7 min you will be listening to a recording through these headphones. Please feel free to adjust the volume as needed, as long as you can still hear the recording clearly. Your only task at this time is to listen to the recording. You will be asked questions about it later. Are there any questions?”

If the participant has questions, they will be answered.

Testing Phase Instructions

“The recording you just heard actually contained several nonsense words. In this task, you are going to hear pairs of syllable-strings. After you have heard both, your job is to indicate which one sounded more like it might be a word from the recording you listened to. If you think the first syllable string sounded more like a word, write ‘1’ on the space provided. If you think the second syllable string sounded more like a word, write ‘2’. You will hear a total of 36 pairs. Do you have any questions?”

APPENDIX D

Questionnaire

Participant Information

Participant # _______

1. How old are you? ____________

2. What is your sex (circle one)? Male Female

3. Is English your native language (circle one)? Yes No

   3a. If no, what is your native language? ______________________________

4. Do you speak any other languages (circle one)? Yes No

   4a. If yes, how many? ____________

   4b. Please list the languages you speak and how many years you have spoken them:

5. Do you have any hearing impairments (circle one)? Yes No

   5a. If yes, do you use assistive devices, such as hearing aids? Yes No

6. Have you ever had any formal musical instruction? Yes No

   6a. If yes, how many years? ______

   6b. If yes, instrumental or vocal? ______________________________

APPENDIX E

Forced-choice Test

Discrimination Task

For each sound pair you hear, please indicate if the first (1) or the second (2) sounds like it might be a word from series of sounds you heard at the beginning of the session by writing the corresponding number in the space provided.

1. __________
2. __________
3. __________
4. __________
5. __________
6. __________
7. __________
8. __________
9. __________
10. __________
11. __________
12. __________
13. __________
14. __________
15. __________
16. __________
17. __________
18. __________
19. __________
20. __________
21. __________
22. __________
23. __________
24. __________
25. __________
26. __________
27. __________
28. __________
29. __________
30. __________
31. __________
32. __________
33. __________
34. __________
35. __________
36. __________

REFERENCES

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9, 4,* 321-324.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1999). Statistical learning in linguistic and nonlinguistic domains. In B. MacWhinney (Ed.), *The Emergence of Language* (359-380). New Jersey: Lawrence Erlbaum Associates Inc.

Becker, J. (2004). *Deep listeners: Music, emotion, and trancing*. Bloomington: Indiana University Press.

Besson, M., & Schön, D. (2000). Comparison between language and music. In I. Peretz & R. Zatorre (Eds.). *The Cognitive Neuroscience of Music* (pp. 269-293). New York, NY: Oxford University Press.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16, 6,* 451-459.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93-125.

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience, 23,* 2791-2803.

Cochran, B. P, McDonald, J. L., & Parault, S. J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language, 41,* 30-58.

Content, A., Kearns, R. K., & Frauenfelder, U. H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language,45, 2,* 177-199.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology, 30, 5,* 1119-1130.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics, 21,* 103-108.

Cutler, A., & Otake, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of Memory and Language, 46, 2,* 296-322.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Verbal Learning and Behavior, 36, 2,* 202-225.

Edelman, G.M. (2006). *Second nature: Brain science and human knowledge*. New Haven, CT: Yale University Press.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8,* 181-195.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language, 32,* 249-268.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18, 3,* 254-260.

Harris, Z. S. (1955). From phoneme to morpheme. *Language, 31, 2,* 190-222.

Hayes, J. R., & Clark, H. H. (1970). Experiments on the segmentation of an artificial speech analogue. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (221-243). New York: John Wiley & Sons Inc.

Hockema, S. A. (2006). Finding words in speech: An investigation of American English. *Language Learning and Development, 2, 2,* 119-146.

Houston, D. M., Santelmann, L. M., & Jusczyk, P. W. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes, 19, 1,* 97-136.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning; The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60-99.

Jusczyk, P. W. (1993). From general to language-specific capacities: the WRAPSA model of how speech perception develops. *Journal of Phonetics, 21,* 3-28.

Jusczyk, P. W. (1997). Finding and remembering words: Some beginnings by English-learning infants. *Current Directions in Psychological Science, 6, 6,* 170-174.

Jusczyk, P. W. (1999). Narrowing the distance to language: One step at a time. *Journal of Communication Disorders, 32,* 207-222.

Jusczyk, P. W. (2002). How infants adapt speech-processing capacities to native-language structures. *Current Directions in Psychological Science, 11, 1,* 15-18.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., & Svenkerud, V. Y. (1993). Infant's sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32, 3,* 402-420.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16,* 55-68.

Kempe, V., Brooks, P. J., Gillis, S., & Samson, G. (2007). Diminutives facilitate word segmentation in natural speech: Cross-linguistic evidence. *Memory & Cognition, 35, 4,* 762-773.

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language, 37,* 295-311.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83,* B35-B42.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,*831-843.

Kuhl, P. K. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255, 5044,* 606-608.

Krentz, U. C., & Corina, D. P. (2008). Preference for language in early infancy: The human language bias is not speech specific. *Developmental Science, 11, 1,* 1-9.

Marcus, G. F., Fernandes, K. J., 7 Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18, 5,* 387-391.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month old infants. *Science, 283, 5398,* 77-80.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38,* 465-494.

McDonald, J. L. (1997). Language acquisition: The acquisition of linguistic structure in normal and special populations. *Annual Review of Psychology, 48,* 215-241.

Mehler, J., Peña, M, Nespor, M., & Bonatti, L. (2006). The “soul” of language does not use statistics: reflections on vowels and consonants. *Cortex, 42, 6,* 846-854.

Mirman, D., Magnuson, J. S., Estes, K. G., Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108,* 271-280.

Morton, J., & Long, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior, 15,* 43-51.

Muise, J. G., Le Blanc, R. S., & Jefferey, C. J. (1972). Comparison of redundancy of French and English through use of a regular finite Markov process. *Perceptual and Motor Skills, 34,* 712-714.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127-162.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology, 49,* 85-117.

Olds Batchelder, E. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition, 83,* 167-206.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53, 2,* 225-237.

Patel, A. D., & Daniele, J. R. (2002). An empirical comparison of rhythm in language and music. *Cognition, 87,* B35-B45.

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception, 24, 1,* 99-104.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298, 5593,* 604-607.

Peretz, I., Radeau, M., & Arguin, M. (2004). Two-way interactions between music and language: Evidence from priming recognition of tune and lyrics in familiar songs. *Memory & Cognition, 32, 1,* 142-152.

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology, 56,* 89-114.

Perruchet, P., & Vinter, A. (1998). PARSER: A model of word segmentation. *Journal of Memory and Language, 39, 2,* 246-263.

Perruchet, P. & Peereman, R. (2003). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics, 17,* 91-119.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20, 2,* 421-435.

Rauscher, F.H. (2008). The impact of music instruction on other skills. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford handbook of music psychology* (pp. 244-252). London: Oxford University Press.

Rauscher, F.H. (2005, October). *Musical influences on children's cognitive development: Effects of quality of instruction*. Royal College of Music, London, UK.

Saffran, J. R. (2001a). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44,* 493-515.

Saffran, J. R. (2001b). Words in a sea of sounds: the output of infant statistical learning. *Cognition, 81,* 149-169.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47,* 172-196.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35, 4,* 606-621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8, 2,* 101-105.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27-52.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4, 2,* 273-284.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39, 3,* 484-494.

Schellenberg, E. G., & Peretz, I. (2007). Music, language, and cognition: unresolved issues. *Trends in Cognitive Science, 12, 2,* 45-46.

Schön, D., Leigh Gordon, R., & Besson, M. (2005). Musical and linguistic processing in song perception. In G. Avanzini, L. Lopez, S. Koelsch, & M. Manjno (Eds.). *The Neurosciences and Music II: From Perception to Performance* (pp. 71-81). New York, NY: New York Academy of Sciences.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language recognition. *Cognition, 106,* 975-983.

Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science, 11, 4,* 596-606.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluid speech. *Cognitive Psychology, 54,* 1-32.

Swingley, D. (2004). Statistical clustering and the contents of infant vocabulary. *Cognitive Psychology, 50,* 86-132.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7-to-9-month-old infants. *Developmental Psychology, 39, 4,* 706-716.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3, 1,* 1-42.

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science, 19, 2,* 137-144.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 91,* B25-B34.

Ushakova, T. N. (2000). Language emergence in infants. *European Psychologist, 5, 4,* 285-292.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9, 4,* 325-329.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to signal: the privileged status of speech for young infants. *Developmental Science, 7, 3,* 270-276.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science, 10, 2,* 159-171.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24, 5,* 672-683.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

Wilson, E.O. (1998). *Consilience: The unity of knowledge*. New York: Knopf.

Yang, C. (2004). Universal grammar, statistics, or both? *Trends in Cognitive Sciences, 8, 10,* 451-456.

Zatorre, R. J., Evans, A. C., Meyers, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256, 5058,* 846-849.

OSHKOSH

December 3, 2008

Ms. Rebecca Horn
529A Central St.
Oshkosh, WI 54901

Dear Ms. Horn:

On behalf of the UW Oshkosh Institutional Review Board for Protection of Human Participants (IRB), I am pleased to inform you that your application has been approved for the following research: Tonal Facilitation of Language Acquisition.

Your research has been categorized as NON-EXEMPT, which means it is subject to compliance with federal regulations and University policy regarding the use of human participants as described in the IRB application material. Your protocol is approved for a period of 12 months from the date of this letter. A new application must be submitted to continue this research beyond the period of approval. In addition, you must retain all records relating to this research for at least three years after the project's completion .

Please note that it is the principal investigator's responsibility to promptly report to the IRB Committee any changes in the research project, whether these changes occur prior to undertaking, or during the research. In addition, if harm or discomfort to anyone becomes apparent during the research, the principal investigator must contact the IRB Committee Chairperson. Harm or discomfort includes, but is not limited to, adverse reactions to psychology experiments, biologics, radioisotopes, labeled drugs, or to medical or other devices used. Please contact me if you have any questions (PH# 920/424-7172 or e-mail: rauscher@uwosh.edu).

Sincerely,

Dr. Frances Rauscher
IRB Chair

INSTITUTIONAL REVIEW BOARD
UNIVERSITY OF WISCONSIN OSHKOSH · 800 ALGOMA BLVD · OSHKOSH WI 54901
(920) 424-3215 · FAX: (920) 424-3221