UNIVERSITY OF CALIFORNIA

Los Angeles

The role of intonation in language discrimination by infants and adults

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Linguistics

by

Chad Joseph Vicenik

2011

The dissertation of Chad Joseph Vicenik is approved.

Toben Mintz

Bruce Hayes

Patricia Keating

Sun-Ah Jun

Megha Sundara, Committee Chair

University of California, Los Angeles

2011

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank Megha for all her help and advice during this project. She was an excellent chair – knowing when to crack the whip and when not to, and for remaining steadfastly optimistic about it all when I was depressed. It could not have been done without her.

Thanks to Sun-Ah, who I would say is my second major linguistic mentor. You've taught me so much and given me so much advice over the years. And to Pat, who has also taught me so much, both directly and indirectly.

I would also like to thank my other committee members, Bruce and Toby. I'm certain I did not take advantage of you guys as much as I probably should have.

Thanks to Adrienne, Kristi, Vicky, Nancy, Yun, Victoria, Stephanie, Spencer, Svetlana, Fiona, Flor, Dulce and all the volunteers in the Language Lab for scheduling, testing and all the hard work you guys do. Thanks to Joseph for all your help in preparing the stimuli and testing adults, you set the standard for all future RAs. I would also like to thank Christina Kitamura, Volker Dellwo, Theirry Nazzi and Constanze Weise for help in recording or providing experimental stimuli.

To Marc, Jamie, Robyn, Craig, Byron, Melanie, Lisa, Anya and all the other members of the department – faculty, students and staff – thanks for being cool. And to Vincent for basically letting me change his name to Sherlock.

Many thanks to Alan Cline for pouring his life into the Dean's Scholars program at UT Austin. You set up an environment that was so supportive and open to exploring

and learning, and Elaine helped make it into a second family. It helped shape who I am today, and we DSers can never thank you enough.

Finally, thank you to my family for all the usual stuff. And to mom – here. It's done. It's filed. Now, can you stop bothering me about it already?

VITA

January 5, 1983	Born, Houston, Texas
2005	B.S., Physics B.S., Astronomy University of Texas, Austin
2008	M.A., Linguistics University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Jun, S., Vicenik, C. & Lofstedt, I. 2007. "Intonational Phonology of Georgian." UCLA Working Papers in Phonetics, 106, pg. 41-57.

--- , Vicenik, C. & Lofstedt, I. 2007. "Intonational Phonology of Georgian." Poster presented at Workshop on Intonational Phonology: Understudied or Fieldwork Languages. A satellite meeting of ICPhS, Saarbruecken, Germany.

--- , Vicenik, C. & I. Lofstedt. 2008. "Intonational Phonology of Georgian", given at the Special Session on Phonetics and Phonology, organized by Donca Steriade, the 18th meeting of *International Congress of Linguists*, Seoul, Korea.

Shue, Y.-L., P. Keating & C. Vicenik. 2009. "VOICESAUCE: A program for voice analysis." Poster at the 158th meeting of the *Acoustical Society of America*, San Antonio, Texas.

Vicenik, C. 2008. "An Acoustic Analysis of Georgian Stops." Unpublished UCLA Masters Thesis.

--- . 2010. "An Acoustic Study of Georgian Stop Consonants." *Journal of the International Phonetic Association* 40(1), p. 59-92.

--- , & Jun, S. (In press). "Intonational Phonology of Georgian," in *Prosodic Typology II: The Phonology of Intonation and Phrasing*, edited by Sun-ah Jun. Oxford University Press.

---, & Kuo, G. 2010. "Tongan Intonation." Poster at the 160th meeting of the *Acoustical Society of America*, Cancun, Mexico.

---, & Sundara, M. 2008. "Role of segmental and intonational cues in dialect discrimination." Poster at Acoustics'08, the 155th meeting of the *Acoustical Society of America*, Paris.

---, & Sundara, M. 2009. "Language and dialect discrimination by five-month-olds." Poster at the 158th meeting of the *Acoustical Society of America*, San Antonio, Texas.

---, & Sundara, M. 2009. "Role of rhythmic and intonational cues in language discrimination." Poster at 2nd Special workshop on speech: Cross-Language Speech Perception and Linguistic Experience, the 157th meeting of the *Acoustical Society of America*, Portland, Oregon.

---, & Sundara, M. 2010. "Language discrimination by infants: discriminating within the native language rhythm class." Poster at 2010 International Conference on Infant Studies, Baltimore, Maryland.

---, & Sundara, M. (Under revision). "The role of rhythm and intonation in language and dialect discrimination by adults." Submitted to *Journal of the Acoustical Society of America.*

ABSTRACT OF THE DISSERTATION

The role of intonation in language discrimination by infants and adults

by

Chad Joseph Vicenik

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2011

Professor Megha Sundara, Chair

It has been widely shown that infants and adults are capable of using only prosodic information to discriminate between languages. However, it remains unclear which aspects of prosody, either rhythm or intonation, listeners attend to for language discrimination. Previous researchers have suggested that rhythm, the duration and timing of speech segments, plays an important role in linguistic processing and drives language discrimination, especially for infants during the first year of life. However, the experiments in this thesis were conducted in an attempt to show that infants and adults strongly attend to intonation, or pitch cues, and, in some cases, weigh this information more heavily than rhythmic information.

The experiments described in chapter 2 test American English-listening adults on their ability to discriminate their native language from a prosodically similar non-native language, German, and from a non-native dialect, Australian English. An acoustic analysis shows there are subtle rhythmic and intonational differences that listeners could potentially use to discriminate between both pairs. Then, American English listeners' ability to discriminate prosodically-similar languages is examined using 1) low-pass filtered and 2) re-synthesized speech, which preserves rhythmic timing and pitch, but degrades or eliminates segmental information, 3) monotone re-synthesized speech, which contains only rhythmic timing information, and 4) scrambled speech, which distorts both rhythmic timing and pitch information. Results show that listeners are capable of using both rhythmic timing and pitch to discriminate between the languages, but weight pitch cues over rhythmic timing cues, even when pitch is uninformative and hinders discrimination.

American English-learning infants' discrimination ability is tested in chapters 3 and 4, using the Headturn Preference Paradigm. In Chapter 3 a series of experiments on infants' discrimination of American English and German are presented. The ability to discriminate between these two languages is found to develop between 5- and 7-months of age. Like adults, 7-month-olds can discriminate between these two languages using only prosodic cues, as evidenced by their ability to discriminate between low-pass filtered speech stimuli. Finally, using stimuli with re-synthesized intonational contours, evidence is provided suggesting that, like adults, infants also weight intonation over other cues such as rhythm.

In Chapter 4 data on American English-learning infants' ability to discriminate of a number of additional language pairs, including American & Australian English, American English & Dutch, British English & Dutch, and American English & Japanese are presented. Infants failed to discriminate these pairs in all but two of the tested cases – five-month-olds were able to discriminate between British English & Dutch and American English & Japanese. However, the behavior observed in these two cases does not fit with previous reports on infant discrimination. The ability to discriminate British English & Dutch was retested using a different methodology, visual habituation, and more typical behavior was observed. The implications for infant testing methodologies are discussed. Finally, in chapter 5, the findings from the current study and concluding remarks are summarized.

1. Introduction

Language identification and discrimination is an important skill that can be taken for granted, considering how easy it is for humans to do it. After hearing only a very small amount of speech, people can accurately identify it as their native language or not, and if not, can often make reasonable guesses about its identity. In fact, this can be done after exposure to only one second of speech (Muthusamy, Barnard & Cole, 1994). The ability to identify and discriminate between languages has obvious importance in basic communication and language acquisition. To the infant, this ability allows for the identification of its native language amidst a potential sea of distracting, yet irrelevant, linguistic noise. For bilinguals (both infants and adults), it allows for the categorization, and discrete learning, of two or more separate input languages. However, for researchers, the ability to discriminate between languages has helped the study of human perceptual abilities and offered insight into how language is processed.

The primary goal of this thesis is to explore the ability of adults and infants to discriminate between their native language and a prosodically similar non-native language. Languages differ in many ways: segmental phonetics, phonology, morphology, syntax and prosody, and each of these can provide cues to a language's identity. However, this thesis will predominantly focus on prosodic cues – rhythm and intonation, specifically – in order to determine which of these is most important to the discrimination task. Previous researchers have suggested that rhythm, the duration and timing of speech segments, plays an important role in linguistic processing and drives language

discrimination, especially for infants during the first year of life. However, in this thesis I will attempt to show that infants and adults strongly attend to intonation, or pitch cues, and, in some cases, weigh this information more heavily than rhythmic information.

In this introductory chapter, I first review research examining language discrimination using unmodified speech. This research clearly illustrates that both adults and infants can discriminate between languages, including non-native language pairs, and that for infants, this ability seems to develop with age. I then discuss three potential cues infants and adults might use to discriminate between languages – segmental cues, rhythmic cues or pitch cues – and previous research examining the roles of these cues. While humans may potentially use any or all of these cues, I suggest that pitch plays a far more important role in language discrimination than previous research suggests.

1.1 Review of language discrimination studies using unmodified speech.

1.1.1 Language discrimination by adults.

It is obvious that adult listeners can discriminate between their native language and a non-native language when presented with unmodified samples of speech. If they are unable to parse any known words or construct any meaning from a continuous speech stream, it is clearly not a language they can speak.

However, adults are also quite capable of discriminating between, and often identifying, two non-native languages. Muthusamy et al. (1994), in an attempt to

establish a benchmark comparison for automated speech recognition technology, played samples from ten different languages to American English-speaking adults and had them identify the language. The stimuli consisted of 1, 2, 4 or 6 s long clips of unmodified speech from English, Farsi, French, German, Japanese, Korean, Mandarin Chinese, Spanish, Tamil and Vietnamese produced by multiple speakers. Even with only 1 s of the speech sample, participants correctly identified the non-native languages 20.7% of the time (chance was 11%). With a full 6 s, accuracy improved to 49.7%.

Muthusamy et al. (1994) only performed a casual examination of the cues used by participants to identify languages. When interviewed, participants reported listening largely to segmental cues, but also to intonation or tone, particularly in the case of French, Mandarin and Vietnamese, and to speech rate in the case of Spanish. Additionally, and unsurprisingly, prior familiarity with a language dramatically increased its identification accuracy.

The facilitating effect that prior familiarity with a language has when discriminating that language from other languages has been shown in other studies, as well. For example, Barkat & Vasilescu (2001) found that Arabic speakers had higher identification accuracy for Arabic dialects from their own geographical region. Western Arabic speakers more reliably identified samples of Western Arabic than Eastern Arabic, while Eastern Arabic speakers showed the opposite pattern. These samples consisted of complete utterances, ranging in duration from 5-30 s, spoken by multiple speakers.

Barkat & Vasilescu (2001) also attempted, in a second experiment, to indirectly determine cues beyond language familiarity that were important in explaining

identification scores by using multi-dimensional scaling (MDS). MDS is a statistical technique that maps a set of stimuli onto a multi-dimensional space based on the similarity ratings for pairs of stimuli. Stimulus items that are viewed as similar to each other are mapped close together, while those deemed dissimilar are mapped further apart. The number of dimensions necessary to adequately describe the similarity space is derived from participants' judgments, and these dimensions are typically correlated with acoustic measures to interpret which cues listeners might use for their judgments. However, in this and the following studies described in this section, researchers made no acoustic measures to test for correlations. Instead, they suggest interpretations, presumably based upon their own perceptual judgments.

Barkat & Vasilescu (2001) tested identification of five Romance languages – French, Spanish, Italian, Portuguese and Romanian, using complete utterances from two speakers per language. These languages were identified by participants from three different language backgrounds – French, Romanian and American English. Barkat & Vasilescu (2001) again found that listeners' identification accuracy depended on their familiarity with the languages. For example, French and Romanian listeners treated their native language differently from non-native languages, while English listeners treated French and Spanish, languages they've likely had some experience with, differently than the three less familiar languages. They also suggest that the segmental properties of the languages were important in discrimination ability. Specifically, they suggest the number of vocalic contrasts used in the languages was important. All three listener groups treated Spanish and Italian, languages with only a front/back vowel distinction,

differently than French, Romanian and Portuguese, languages with a third vowel contrast (e.g., non-peripheral vowels or nasalized vowels).

Other researchers, using results from MDS studies, have suggested additional cues that affect language identification. Stockmal, Muljani & Bond (1996) had American English listeners provide same-different judgments and similarity ratings for samples from Arabic, Mandarin, Indonesian, Japanese, Russian and Spanish. These samples consisted of complete utterances, about 5 s in length, spoken by two speakers per language. Bond, Fucci, Stockmal & McColl (1998) had American English listeners provide similarity ratings for Akan, Arabic, Chinese, English, French, German, Hebrew, Japanese, Latvian, Russian and Swahili. The samples used in this experiment consisted of 10 s samples from a single male speaker from each language. And, Nazzi (1997) had French listeners provide similarity ratings to samples from Arabic, Dutch, English, French, Italian, Japanese, Spanish, Swedish, Turkish, and Wolof, spoken by 4 female speakers of each language.

Nazzi (1997) and Bond et al. (1998) both found an apparent effect of prior familiarity with the languages, while Stockmal et al. (1996) did not discuss the possibility. However, beyond familiarity, the suggested important cues influencing perceived language similarity include speaker voice quality, speech rate (Stockmal et al., 1996), and speaker reading style (either dramatic or formal; Bond et al., 1998). In addition, the researchers in all three reports suggested broadly defined prosodic properties of the languages, such as linguistic rhythm, the role of pitch or the placement of stress, played an important role.

However, there is at least one confound in each of these studies. The samples of the different languages used were each recorded by one or more unique speakers. Thus, it is possible that listeners in these studies were discriminating between speakers rather than between languages. In an attempt to remove effects of speaker identity, Stockmal, Moates & Bond (2000) and Stockmal & Bond (2002) had American English listeners provide same/different responses to paired samples of languages spoken by the same bilingual speakers and found that discrimination was still possible. These pairs used in the earlier study are reprinted in Table 1.1. Stockmal & Bond (2002) also attempted to remove effects of language familiarity by using samples from several prosodically-similar African languages – Akan, Haya, Kikuyu and Luhya – which were each paired with Swahili. As with the earlier studies, they did not provide any accompanying acoustic correlates, but using an MDS analysis, claimed that listeners' similarity ratings were driven by basic prosodic properties of the languages, such as linguistic rhythm or stress placement (Stockmal et al., 2000), or by segmental differences (because all the African languages were deemed prosodically similar; Stockmal & Bond, 2002).

Male talkers	Female talkers
Arabic – French (Morocco)	Korean – Japanese (Asia)
Hebrew – German (Israel)	Ombawa – French (Cameroon)
Akan – Swahili (Africa)	Latvian – Russian (Europe)
Latvian – Russian (Europe)	Ilocano – Tagalog (Philippines)

Table 1.1. Language pairs used in Stockmal et al. (2000), reprinted. In each pair, the talker's primary language is given first.

In only one recent MDS study did researchers attempt to find correlations between acoustic measures and the MDS dimensions. Bradlow, Clopper, Smiljanic &

Walter (2010) had American English-listeners classify 1.5-2 s long samples of speech from several languages, produced by a single male speaker per language. These languages included Amharic, Arabic, Hebrew, Persian, Cantonese, Sindhi, Japanese, Croatian, Catalan, Galician, Hausa, Korean, Dutch, Swedish, Hungarian, Turkish, and Slovene. Their acoustic measures included several speech rate measures, rhythm measures, pitch measures, and a few categorical segment inventory measures, such as whether or not the language had marked velar or dorsal consonants. However, none of the acoustic measures they made had a significant correlation with either of the two MDS dimensions. Rather, they found a tendency for one dimension to correlate with the presence or absence of marked velar segments and a tendency for one dimension to correlate with the presence or absence of marked front round vowels, suggesting their subjects might have classified the languages based on segmental information.

These studies clearly show that adults are quite capable of discriminating between, and even identifying, non-native languages. Further, multi-dimensional scaling has been used in much of this research to determine the similarity among different non-native languages. However, because there were no acoustic measurements made in association with all but one of these studies, the interpretations assigned to the dimensions output by the MDS analyses cannot be considered seriously. As a result, beyond the possibility of segmental information, it remains very unclear what cues adults are using for language discrimination.

1.1.2 Language discrimination by infants – direct evidence

Although the ability of adults to identify languages has been examined in a number of studies, language discrimination has arguably received more focus in infant research. Researchers have looked at language discrimination abilities of infants at different ages and with varied language backgrounds, using different methodologies, for both native and non-native language pairs. This section reviews research that directly tests infants' ability to discriminate between languages, while the following section reviews research that indirectly tests infants' discrimination ability by measuring language preference. The summary of this research outlines a course of development in language discrimination ability over the first year of life.

The first two studies to look at language discrimination by infants were Mehler, Juzcyk, Lambertz, Halstead, Bertoncini & Amiel-Tison (1988) and Bahrick & Pickens (1988). Bahrick & Pickens (1988) tested American English-learning 5-month-olds' discrimination of American English and Spanish. They used a visual habituation procedure (VisHab). In VisHab, infants are seated in front of a television screen. Their attention is attracted to the screen by some method, at which point an audio stimulus plays from a loudspeaker situated below the screen while some, usually unrelated, visual display is projected on the screen. The amount of time the infant fixates on the screen while the sound is played is measured as an indicator of the infant's attention to the stimuli. Bahrick & Pickens (1988) used a bimodal study – that is, the visual display was

relevant to audio stimuli. They played a video of the speaker as she spoke the sentences played over the speaker. Only a single speaker was used in the study.

Infants were exposed to a passage from one of the two languages until the average visual fixation time for the last two trials reduced by a threshold criterion compared to the first two trials, in this case, 50%. They then heard and saw, depending on the experimental group, (1) a new passage in the novel language, (2) a new passage in the familiar language or (3) a previously heard passage. If infants could discriminate between the language samples, then their fixation time should increase when the novel language was presented, indicating dishabituation. If the habituated language was heard (regardless of whether the passage was novel or familiar), their fixation time should remain unchanged. They found evidence of language discrimination using the audio-visual stimuli – American English-learning 5-month-olds only dishabituated to the novel language sample. A replication of the experiment without audio cues showed no evidence of discrimination, indicating that infants required the audio to discriminate between the languages. No experiment testing audio cues only was conducted.

Mehler et al. (1988), on the other hand, performed an entirely audio-based study. In that study, 4-day-old monolingual French-learning infants and 2-month-old American English-learning infants were tested on their ability to discriminate between two language pairs: French & Russian and American English & Italian, or in other words, a pair containing their native language and a pair of two non-native languages. This study used two different testing methodologies for the two age groups. For the newborns, they used

the high-amplitude sucking procedure (HAS; Eimas et al., 1971), and for the 2-month-olds, they used VisHab, both in a habituation-change set up.

In a HAS study, infants suck on a pacifier connected to a pressure transducer that allows a computer to track the sucking rate of the infant while listening to auditory stimuli. Normally, the audio plays contingent to the infants sucking, however, in this experiment, the audio stimuli were played regardless of the sucking behavior. Infants were habituated to samples of speech from one language, which were composed of groups of sentences lasting between 13-22 s, spoken by a single speaker. The rate of sucking was measured for each speech sample. An infant was considered to have habituated to the language when the sucking rate, averaged over the last three stimulus items, had fallen by a threshold amount compared to the first three stimulus items. In this case, the habituation threshold was a decline of 33%. Once infants were habituated, they were moved into a test phase. Depending on the test group, infants either heard 9 new stimulus items in the same language or 9 new items in a new language. The same bilingual speaker was used to create all the stimuli.

In the VisHab study, like in the HAS study, infants were habituated to stimuli from a single language until their average fixation time for the last three stimulus items had fallen by a threshold amount compared to the average fixation time for the first three items. For the VisHab experiments, this threshold was 50%. Depending on the test group, when they were habituated, infants either heard 3 new stimulus items in the same language or 3 new items in a new language. The same bilingual speaker was used to create all the stimuli.

In both methods, if infants could discriminate between the language samples, then their sucking rate or fixation time should increase when the novel language was presented, indicating dishabituation. If the familiar/habituated language was heard, their sucking rate/fixation time should remain unchanged. Mehler et al. (1988) initially found that both newborns and 2-month-old infants only showed discrimination on the language pairs that contained their native language. That is, French newborns could discriminate between French & Russian and American English-learning 2-month-olds could discriminate between American English & Italian. Discrimination was not observed for the non-native language pair. For the newborns, there was also an observed asymmetry – French newborns only dishabituated when switching from Russian to French. Newborns habituated to French and then tested on Russian did not show a significant dishabituation. They also tested newborns from a variety of language backgrounds, none of them French or Russian. These non-French/Russian-learning newborns were also unable to discriminate between French & Russian. These results led Mehler et al. to posit that newborns had already homed into their native language at birth, and that language discrimination was contingent on recognizing the native language.

However, these data were re-examined in Mehler & Christophe (1995), resulting in a different conclusion. In the original study, Mehler et al. (1988) kept groups separated by the habituation language. For example, newborns habituated to French and then tested on Russian were analyzed separately from newborns habituated to Russian and then tested on French. Typically, these groups are combined and only the difference in behavior between habituated and novel languages is examined.

When the data were reanalyzed in this more typical way, it was found that French-learning newborns could discriminate between both French & Russian and American English & Italian, suggesting that discrimination was not simply based on recognizing the native language. Rather, newborns could discriminate both their native language from a non-native language, as well as two non-native languages. However, American English-learning 2-month-olds were still unable to discriminate between the two non-native languages French & Russian. Mehler & Christophe interpreted this as support for a possible developmental track – that while younger infants attended to all the utterances to which they are exposed, by 2-months, they already concentrate only on the utterances of their native language and neglect those from a non-native language (Mehler & Christophe, 1995). However, it is worth noting that, contrary to this hypothesis, in Mehler et al's (1988) study, the non-French/Russian-learning newborns showed no evidence of an ability to discriminate between French & Russian.

Some confirmatory evidence for a developmental trajectory was found in another study by Christophe & Morton (1998). They tested British English-learning 2-month-olds' discrimination of English & Japanese, French & Japanese, English & Dutch and Dutch & Japanese using the high-amplitude sucking procedure in a habituation-change setup. Unlike Mehler et al's (1988) between-subjects design, Christophe & Morton (1998) used a within-subjects design – every infant was tested on the familiar as well as the change languages. Their stimuli consisted of sentences recorded by four native speakers of each language in the respective pairs. British English-learning 2-month-olds were able to discriminate English & Japanese, but not French & Japanese. These results

support the notion that by 2-months of age, infants have started to home in on their native language and are less sensitive to differences in non-native languages.

Christophe & Morton (1998) report only equivocal evidence for discrimination between English & Dutch and Dutch & Japanese. Although a majority of infants had a higher sucking rate for the novel language than for the habituated language, the result was only marginally significant in both cases. The behavior of infants in these groups was also not significantly different from the infants tested on English & Japanese - they showed clear evidence of discrimination - and only marginally different from the infants tested on French & Japanese - they failed to show discrimination. If infants had honed in on their native language by this age, they should have discriminated between English & Dutch, but not Dutch & Japanese. However, English and Dutch are historically related languages that share a number of features. Thus, perhaps British English-learning 2-month-olds were unable to distinguish Dutch from English because of the similarity of the two languages.

This was confirmed by Nazzi, Jusczyk & Johnson (2000). Nazzi et al. (2000) looked at the discrimination ability of an older group of infants – American English-learning 5-month-olds. They used the Headturn Preference Parardigm (Kemler Nelson, Jusczyk, Mandel, Myers, Turk & Gerkin, 1995) with a familiarization-novelty setup, a within-subjects design, and stimuli from multiple talkers. In this procedure, infants sit on a caregiver's lap in a three-sided booth, with a light on each wall. The lights are used to capture the infant's attention. Once they are looking at a side light, sound from a uni-directional speaker hidden behind the light plays audio stimuli. Infants were exposed to

stimuli from a single language for a predetermined amount of time. Unlike a habituation-change setup, infants did not have to meet a criterion showing a decline in fixation time. Rather, infants simply had to hear a minimum of 80 s of stimulus sentences presented in one language – 20 s each of four passages spoken by two speakers (two passages per speaker). After the familiarization phase, infants were exposed to new language samples, spoken by previously unheard speakers, in previously familiarized language and a novel language. Like in a habituation-change study, if infants successfully discriminate between the language samples, then their fixation time should be longer for the novel language than for the familiar language. That is, infants are expected to show a novelty effect.

Nazzi et al. (2000) tested 5-month-olds on a number of language pairs: British English & Japanese, Spanish & Italian, British English & Dutch, German & Dutch, and American & British English.[1] Consistent with Christophe & Morton's findings, American English-learning 5-month-olds were able to discriminate their native language from Japanese, but failed to discriminate between two non-native languages, either Spanish-Italian or German-Dutch. In contrast to the marginal results from Christophe & Morton (1998), 5-month-olds showed clear evidence of discrimination between British English & Dutch, and were even able to discriminate between two dialects of their language, American & British English. Both Christophe & Morton (1998) and Nazzi et al. (2000), unlike the previous studies, use stimuli from multiple speakers. This allows

[1] They also tested Italian & Japanese, which will be discussed in section 1.2.2

for greater generalization from their results – it is clear that the observed effects are not caused by the specific talker.

All these studies together show that the ability to discriminate between languages develops over the first year of life. At birth, infants are sensitive to and can discriminate between some language pairs which may not include their native language. For example, French-newborns can discriminate between English and Russian. Yet, as they age, they hone in on their native language, resulting in a declining ability to discriminate between non-native languages, and an increasing ability to discriminate their native language from non-native languages.

1.1.3 Language discrimination by infants – indirect evidence through language preference.

Studies that have looked at infants' language preference also bear on the question of infant language discrimination because in order to prefer one language over the other, the infant must first discriminate between them. Moon, Cooper & Fifer (1993) tested 2-month-old English-learning and Spanish-learning infants' preference for English versus Spanish. They used a HAS task, though in a preference paradigm. Infants were not habituated to one language, but rather heard samples from both languages, interleaved. Infants should suck more for one language over the other if they show a preference. They used stimuli from multiple speakers - eight female English speakers and eight female Spanish speakers, although any given infant heard only two speakers, one in each

language. Infants preferred to listen to language samples from their native language, indicating that they could discriminate between the two languages. However, this preference was observed only during the last third of the experiment, after infants had heard 12 min of alternating stimuli.

Deheane-Lambertz & Houston (1997) tested 2-month-old French-learning and English-learning infants on French and English using a visual orientation latency paradigm (VOLP). In VOLP, infants' attention is drawn to a central monitor by quiet visual stimuli. Then, language stimuli are played on one of two loudspeakers – one on either side of the central monitor. The time between the onset of audio stimuli and an infant's orientation towards the source is measured. The paradigm is based on the hypothesis that if the two languages being presented are processed differently, then the orientation time for one should be faster than the other. Different orientation times to two stimuli also indicate they can be discriminated. Typically, this is caused by a familiarity preference – infants orient faster to things they are familiar with, at least based on developmental studies using vision stimuli (Johnson et al., 1991; Schonen et al., 1993).

In their study, Dehaene-Lambertz & Houston (1997) used stimuli recorded by a single bilingual female speaker. They found that English-learning infants oriented faster to their native language than to French, while French-learning infants showed no difference in orientation times to the two languages. If infants showed a familiarity preference, as is typically expected in VOLP tasks, then French infants should have oriented faster to French. They suggest that these results are due to a combination of a cross-linguistic preference for English and a familiarity preference for the native

langauge. For English-learning 2-month-olds, these two effects work in the same direction, but for French-learning 2-month-olds, they cancel each other out. Dehaene-Lambertz & Houston (1997) seem to assume that both English-learning and French-learning infants can discriminate between the two languages. However, it is possible that French-learning infants did not show a language preference because they could not discriminate between them.

Differences in preference by infants from different language backgrounds have been seen in another study. Byers-Heinlein, Burns & Werker (2010) tested Canadian English monolingual, English-Tagalog bilingual, and English-Chinese bilingual newborns (both English-Mandarin and English-Cantonese) on their preference for Canadian English or Tagalog using a HAS preference task. They used stimuli recorded from multiple speakers, but low-pass filtered at 400 Hz (see section 1.2.1). They found that English monolingual newborns had a higher sucking rate while listening to stimuli from their native language than to Tagalog stimuli. Neither bilingual group showed a significant preference for either language.

Byers-Heinlein et al. (2010) did test to see if this lack of preference was due to an inability to discriminate between the languages. Using a HAS habituation-change discrimination task, they observed that English-Tagalog bilingual newborns could indeed distinguish the two languages. Thus, it is clear that a lack of language preference cannot be taken as evidence for a lack of discrimination. In this case, the lack of preference by bilinguals seems to result from their linguistic background.

Bosch & Sebastián-Gallés (1997) also found that bilingual infants showed different preference behavior than monolinguals. They tested 4-month-old Spanish-learning, Catalan-learning and bilingual Spanish-Catalan-learning infants in a number of VOLP experiments. They tested different pairings of Spanish, Catalan, English and Italian, using stimuli recorded from a single quadralingual speaker. For monolingual learners, they found that infants could discriminate between their native language (Spanish or Catalan) and a non-native language (English, and Spanish or Catalan). As expected with this paradigm (Johnson et al., 1991; Schonen et al., 1993), infants oriented faster towards the familiar, native langauge.

In contrast, bilingual Spanish-Catalan infants were tested on their maternal language and English or Italian, and discrimination was observed for both language pairs. In this case, infants oriented faster towards the non-familiar langauge. When tested on a pair consisting of both their native languages, bilingual infants showed equal orientation times for Spanish and Catalan. Bosch & Sebastián-Gallés suggest that this may be because VOLP actually tests recognition, rather than familiarity preference. Because bilinguals know two languages, it takes them longer to identify which of the two familiar languages they're hearing. Indeed, their orientation time to the unfamiliar language (English or Italian) was not significantly different from monolinguals.

This explanation is reasonable, but regardless of its accuracy, it is clear that bilinguals were able to discriminate between the tested languages. In a later study, Bosch & Sebastián-Gallés (2001) replicated these results using a HPP familiarization-novelty task. Monolingual Spanish and Catalan-learning and bilingual 4-month-olds were able to

discriminate between Spanish and Catalan. In this paradigm, there was no difference between the groups. Both showed the expected novelty preference.

Hayashi, Tamekawa, Kiritani (2001) looked at the development of language preference by conducting a study on Japanese-learning infants from a wide range of ages using an HPP task. Unlike the familiarity-novelty setup used by Nazzi et al. (2000), Hayashi et al. (2001) used a basic preference paradigm. Like in Moon et al. (1993) and Byers-Heinlein et al. (2010), infants were not familiarized to any one language, but rather were played stimuli from the two languages interleaved with each other. They used sentences in English and Japanese recorded by one bilingual speaker, spoken in an infant-directed register to test infants between 4- and 14-months of age. In their analysis, they divided the infants into three age groups, 4-6-months, 7-9-months and 10-14-months. They found that infants showed no significant preference to either language until 7- to 9-months. Infants at this age and older preferred to listen to their native language, Japanese. They also found that the magnitude of this preference increased with age. In a similar study, Sundara, Kitamura & Nazzi (2008) found that Australian English-learning infants also preferred their native language over Japanese, and that this preference increased with age.

Each of these studies indicates that monolingual-learning infants, when they show a preference at all, prefer to listen to their native language over a non-native language. However, the lack of preference by 4-6-month-olds found in Hayashi et al. (2001) and Sundara et al. (2008) contrasts with Byers-Heinlein et al. (2010) and Moon et al.'s (1993) studies because the latter found evidence of a language preference at birth and 2-months,

respectively. Byers-Heinlein et al. (2010) found evidence of a preference during the first 10 minutes of language exposure, but Moon et al. (1993) found that 2-month-olds did not show a preference until after they had already been exposed to 12 minutes of stimuli. Hayashi et al. (2001) not only used a different methodology (HPP vs. HAS), but a shorter procedure, consisting of only 16 total trials. At an average listening time of ~10 s per trial, infants in their experiment would have only heard around 3 minutes of stimuli.

It is difficult to interpret these varying results – the presence or lack of preference, and when it may or may not occur – because these studies differ in many ways. For example, it may be possible that if the procedure used in Hayashi et al. (2001) was longer, younger infants may have shown a native language preference. However, it is possible that infants show different responses to different language pairs. The English and Tagalog stimuli used by Byers-Heinlein et al. (2010) may simply be easier to distinguish than the stimuli used in the other studies, allowing infants to show a native language preference at an earlier age, or after listening to less stimuli.

As a whole, these studies serve to add a number of new language pairs to the infant discrimination literature, and show that, for monolingual infants, when a language preference is observed, it is a preference for the native language. They also illustrate that how preference data relate to language discrimination is not always predictable or easily interpretable.

Language preference research also supports the idea that, as they age, infants' ability to discriminate their native language from a non-native language increases, specifically the research by Bosch & Sebastián-Gallés (1997). Recall that 2-month-old

English-learning infants cannot discriminate between their native language and a non-native language, English and Dutch (Christophe and Morton, 1998), but 5-month-old English-learning infants can (Nazzi et al., 2000). In contrast, 5-month-old English-learning infants cannot discriminate between Spanish and Italian, two non-native languages (Nazzi et al., 2000), but 4-month-old Spanish-Catalan bilingual infants can, since one of the pair is their native language (Bosch & Sebastián-Gallés, 1997). Thus, discrimination between related or prosodically similar languages by older infants depends on one of the languages being their native language.

1.2 Cues for language discrimination.

The previous two sections showed that the ability to discriminate languages develops with age. Infants are born with the general ability to discriminate between some languages, including some non-native language pairs. Recall that French newborns were able to discriminate English & Italian (Christophe & Mehler, 1995). Over the next several months, they begin to home in on their native language. This perceptual narrowing results in a decline in the ability to discriminate non-native languages. For example, English-learning 2-month-olds cannot discriminate between French-Russian (Mehler et al., 1988) or French-Japanese (Christophe & Morton, 1998). However, this homing also allows for the discrimination of previously indistinguishable languages. For example, although English-learning 2-month olds cannot discriminate between English-

Infants' Native Language	Age	Tested Languages			Non-native Pair?	Evidence of Discrim.	Stimuli	Num. Speakers	Method	Setup	Habit-uation Criterion	Source
Canadian English	newborn	Canadian English	&	Tagalog		Yes	Low-Pass	Multi	HAS	Hab-Chg	25%	Byers-Heinlein et al., 2010
Canadian Eng.-Tagalog biling.	newborn	Canadian English	&	Tagalog		Yes	Low-Pass	Multi	HAS	Hab-Chg	25%	Byers-Heinlein et al., 2010
French	newborn	Italian	&	American English	Yes	Yes	Full Cue	Single	HAS	Hab-Chg	33%	Mehler & Christophe 1995
French	newborn	French	&	Russian		Yes	LP & FC	Single	HAS	Hab-Chg	33%	Mehler et al. 1988
Mixed	newborn	French	&	Russian	Yes	No	Full Cue	Single	HAS	Hab-Chg	33%	Mehler et al. 1988
American English	newborn	American English	&	Spanish		Yes	Full Cue	Two	HAS	Pref		Moon, Cooper & Fifer 1993
Spanish	newborn	Spanish	&	American English		Yes	Full Cue	Two	HAS	Pref		Moon, Cooper & Fifer 1993
French	newborn	Dutch	&	British English	Yes	No	Low-Pass	Multi	HAS	Hab-Chg	25%	Nazzi, Bertoncini & Mehler 1998
French	newborn	Japanese	&	British English	Yes	Yes	Low-Pass	Multi	HAS	Hab-Chg	25%	Nazzi, Bertoncini & Mehler 1998
French	newborn	Dutch	&	Japanese	Yes	No	Full Cue	Multi	HAS	Hab-Chg	25%	Ramus, 2002
French	newborn	Dutch	&	Japanese	Yes	Yes	Resynth	(Single)	HAS	Hab-Chg	25%	Ramus, 2002
British English	2-months	British English	&	Japanese		Yes	Full Cue	Multi	HAS	Hab-Chg	20%	Christophe & Morton 1998
British English	2-months	Dutch	&	Japanese	Yes	No	Full Cue	Multi	HAS	Hab-Chg	20%	Christophe & Morton 1998
British English	2-months	British English	&	Dutch		No	Full Cue	Multi	HAS	Hab-Chg	20%	Christophe & Morton 1998
British English	2-months	French	&	Japanese	Yes	No	Full Cue	Multi	HAS	Hab-Chg	20%	Christophe & Morton 1998
American English	2-months	American English	&	French		Yes	LP & FC	Single	VOLP			Dehaene-Lambertz & Houston 1998
French	2-months	American English	&	French		No	LP & FC	Single	VOLP			Dehaene-Lambertz & Houston 1998
American English	2-months	American English	&	Italian		Yes	LP & FC	Single	VisFix	Hab-Chg	50%	Mehler et al. 1988
American English	2-months	French	&	Russian	Yes	No	Full Cue	Single	VisFix	Hab-Chg	50%	Mehler et al. 1988
Catalan	4-months	Spanish	&	Catalan		Yes	LP & FC	Single	VOLP			Bosch & Sebastian-Galles, 1997
Spanish	4-months	Spanish	&	Catalan		Yes	LP & FC	Single	VOLP			Bosch & Sebastian-Galles, 1997
Catalan	4-months	Catalan	&	American English		Yes	Full Cue	Single	VOLP			Bosch & Sebastian-Galles, 1997
Spanish	4-months	Spanish	&	American English		Yes	Full Cue	Single	VOLP			Bosch & Sebastian-Galles, 1997
Spanish-Catalan bilingual	4-months	Maternal language	&	American English		Yes	Full Cue	Single	VOLP			Bosch & Sebastian-Galles, 1997
Spanish-Catalan bilingual	4-months	Maternal language	&	Italian		Yes	Full Cue	Single	VOLP			Bosch & Sebastian-Galles, 1997
American English	5-months	American English	&	Spanish		Yes	Full Cue	Single	VisFix	Hab-Chg	50%	Bahrick and Pickens, 1988
American English	5-months	British English	&	Japanese		Yes	Full Cue	Multi	HPP	Fam-Nov		Nazzi, Jusczyk & Johnson, 2000
American English	5-months	Italian	&	Japanese	Yes	Yes	Full Cue	Multi	HPP	Fam-Nov		Nazzi, Jusczyk & Johnson, 2000
American English	5-months	Spanish	&	Italian	Yes	No	Full Cue	Multi	HPP	Fam-Nov		Nazzi, Jusczyk & Johnson, 2000
American English	5-months	British English	&	Dutch		Yes	Full Cue	Multi	HPP	Fam-Nov		Nazzi, Jusczyk & Johnson, 2000
American English	5-months	German	&	Dutch	Yes	No	Full Cue	Multi	HPP	Fam-Nov		Nazzi, Jusczyk & Johnson, 2000
American English	5-months	American English	&	British English	Yes	Yes	Full Cue	Multi	HPP	Fam-Nov		Nazzi, Jusczyk & Johnson, 2000
S.W. British English	5 & 7-months	S.W.British English	&	Welsh English		Yes	Full Cue	Multi	HPP	Fam-Nov		Butler et al., 2010
S.W. British English	5-months	Scottish English	&	Welsh English	Yes	No	Full Cue	Multi	HPP	Fam-Nov		Butler et al., 2010
Australian English	3-16-months	Australian English	&	Japanese		Yes	Full Cue	?	VisFix	Pref		Sundara et al. 2008
Australian English	3-16-months	Australian English	&	German		No	Full Cue	?	VisFix	Pref		Sundara et al. 2008
Japanese	4-14-months	Japanese	&	?? English		Yes	Full Cue	Single	HPP	Pref		Hayashi et al. 2001
French	newborn	British-Dutch	&	Spanish-Italian	Yes	Yes	Low-Pass	Multi	HAS	Hab-Chg	25%	Nazzi, Bertoncini & Mehler 1998
French	newborn	British-Spanish	&	Dutch-Italian	Yes	No	Low-Pass	Multi	HAS	Hab-Chg	25%	Nazzi, Bertoncini & Mehler 1998

Table 1.2. A summary of the language pairs that have been tested on infants in discrimination/preference studies, including language background of participants, methodology and results.

Dutch (Christophe & Morton, 1998), 5-month-olds can (Nazzi et al., 2000). Table 1.2 summarizes the language pairs that have been tested on infants in discrimination/preference studies.

What information in the speech signal are infants attending to when discriminating languages, and what have older infants learned about language that allows them to discriminate languages that they were previously unable to distinguish? The VOLP experiments by Deheane-Lambertz & Houston (1998) and Bosch & Sebastián-Gallés (1997) shed some light on the time scale in which the relevant acoustic information is conveyed. In those two studies, infants oriented to the proper side after about 1.5-2 s. Languages tend to be spoken at rates of only 5-8 syllables per second (Dauer, 1983). Thus, it seems that infants can discriminate between languages after hearing only about a dozen syllables. This is true for adults, as well: adults can also discriminate between, and even identify, languages, after as little as 1 s of speech (Muthusamy et al., 1994).

In the following three sections, I will review general acoustic cues listeners could use to discriminate between languages. The first section suggests the use of segmental cues. However, it is unlikely listeners are relying solely on segmental information to discriminate languages because a number of studies have found evidence of discrimination, by both infants and adults, using stimuli with degraded or absent segmental information. The two sections that follow review the possibility that listeners use an aspect of prosodic information – either rhythm or intonation.

1.2.1 Segmental cues or prosodic cues?

One possibility is that infants discriminate between languages using segmental information. It is well known that infants, even newborns, are capable of perceiving a number of phonetic contrasts, including both consonantal and vocalic contrasts, regardless of whether or not the contrasts are in their native language (Eimas et al., 1971; Morse, 1972; Trehub, 1973; 1976; Werker et al., 1981; Werker and Tees, 1984). Two-month-olds are even capable of perceiving allophonic contrasts, like the presence or absence of aspiration or voicing (Hohne and Jusczyk, 1994).

Infants have also been shown to attend to distributional information in language. They can track the transitional probabilities between syllables and use this information to segment word forms from continuous speech, as early as 7-months (Thiessen and Saffran, 2003). They can also track how often segments appear, and presumably calculate which phonemic contrasts are used in a language, as early as 6-months (Maye et al., 2002). With these two abilities, infants could perceive the different segments used in two different language samples, track their distributional frequency, and use that information to discriminate between them. Such a strategy has been suggested by Bosch & Sebastián-Gallés (2001).

However, it seems unlikely that infants are relying on segmental information to discriminate between languages. Although infants are born with the ability to distinguish between non-native phonetic contrasts, this ability reduces with age. Evidence suggests the change in ability to perceive segmental contrasts is due to their growing distributional

awareness of the segments and contrasts important to their native language (Maye et al., 2002; Maye et al., 2008; Yoshida et al., 2010). Although this fits with the strategy given above – that infants track statistical information about the prevalence of different segments in their native language and use this information to discriminate languages – this perceptual reorganization of consonantal contrasts does not occur until 8- to 10-months of age, (Werker and Tees, 1984; Kuhl et al., 1992; Polka and Werker, 1994). However, there is some evidence that the perception of vowels is reorganized by language experience earlier in development, specifically, between 4- and 6-months (Kuhl *et al.*, 1992; Polka and Werker, 1994; Polka and Bohn, 1996), which is the age at which infants' abilities to discriminate languages has been tested in some studies (Bosch & Sebastián-Gallés, 1997; 2001, Nazzi et al., 2000). Therefore, the importance of segmental information cannot be discarded by this reasoning alone.

The fact that in several studies discrimination has been successfully observed when infants were presented with low-pass filtered sentences offers more direct evidence that infants do not need segmental information to discriminate languages. Low-pass filtering eliminates most of the segmental information in speech stimuli while preserving its prosody. Many of the previously mentioned studies extended their initial findings using experiments with low-pass filtered speech. Mehler et al. (1988) found that French newborns could still discriminate French & Russian, and English-learning 2-month-olds could still discriminate English & Italian even when the stimuli had been filtered. Similarly, both Dehaene-Lambertz & Houston (1997) and Bosch & Sebastián-Gallés

(1997) found that discrimination was unchanged by low-pass filtering the stimuli used in their studies as well.

Additional researchers have looked at discrimination using only filtered stimuli. Nazzi, Bertoncini & Mehler (1998) used a HAS habituation-change paradigm (with a habituation criterion of 25%) to test French newborns' discrimination of British English & Japanese and English & Dutch. Unlike Mehler et al. (1988) who used sentences recorded by a single speaker, Nazzi et al. (1998) used sentences from four speakers per language. Infants were habituated to two speakers and then tested on an additional two. They found that French newborns could discriminate between English & Japanese, but not English & Dutch. Similarly, Byers-Heinlein et al. (2010) tested Canadian English monolingual and English-Tagalog bilingual newborns on their ability to discriminate between low-pass filtered samples of English & Tagalog, and found that both groups could distinguish the two languages.

Although experiments using low-pass filtered stimuli provide some evidence to rule out the role of segmental information in predicting language discrimination performance, low-pass filtering does not remove all segmental information – though whatever information survives the filtering process is highly degraded. For example, the first formant of vowels may still be discernable in filtered speech.

The reliance on prosody in language discrimination has been demonstrated in other ways. Besides low-pass filtering, Dehaene-Lambertz & Houston (1998) also tested for discrimination between English & French using scrambled sentences. Scrambling disrupts the prosodic coherence of the overall sentence, but preserves segmental

information and word-level prosody. They found a non-significant trend for both infant groups (French and English-learning 2-month-olds) to orient faster to scrambled French stimuli, but take the lack of significance as evidence for a lack of discrimination. Thus, even though segmental information remained intact, infants could no longer discriminate the scrambled stimuli. In addition to ruling out the role of segmental information, this result also suggests that word-level prosody is not sufficient to cue language discrimination.

A different technique was used by Ramus (2002) to examine the use of prosody in language discrimination. He tested French newborns on re-synthesized Dutch & Japanese sentences. He replaced natural phonemes with re-synthesized phonemes, and then copied the original intonation contours on to the re-synthesized sentences. All stops were replaced with /t/, all fricatives were replaced with /s/, sonorants with /l/, off glides with /j/ and vowels with /a/. This method of re-synthesis, called *saltanaj* speech, is described in Ramus & Mehler (1999). It removes all segmental information, but preserves prosodic information and some phonotactic information about how different segment types are combined. He tested infants using the HAS paradigm in a habituation-change setup with a habituation criterion of 25%, and found that French newborns could still discriminate between Dutch & Japanese.

Based on these studies, it seems unlikely that infants are relying on segmental information to discriminate languages. Instead, infants seem to rely on utterance-level prosodic information to discriminate languages. Most of the adult studies on language discrimination also examine the use of prosody, and show that the ability to use prosodic

cues to discriminate languages carries into adulthood. Bush (1967) and Richardson (1973) used low-pass filtered speech to examine adults' language discrimination abilities. Bush (1967) reports that English speakers were able to identify low-pass filtered sentences from different dialects of English – American, British and Indian English – with high accuracy rates (near 80%). Similarly, Richardson (1973) tested American English-speakers on standard American and African-American dialects of English and also found high accuracy rates, near 73%.

Other techniques like the use of laryngograph signals, sinusoidal or pulse train re-synthesis, and Inverse-LPC-filtering also effectively isolate prosody from segmental information, either by filtering away high frequency information or by replicating prosodic information in new, synthesized sound files. Maidment (1976, 1983) and Moftah & Roach (1988) both used laryngograph signals. This is a measure of glottal electrical resistance and is closely related to a glottal waveform, which is the output of the voice source before it is filtered through the vocal tract. Maidment (1976, 1983) tested discrimination between English-French and Moftah & Roach (1988) tested discrimination of Arabic-English. Moftah & Roach also used low-pass filtered speech in their study. The background of participants was not reported in these short reports. In all these studies, subjects had to attempt to identify the language samples, and evidence of discrimination was found in all cases. Maidment found higher language identification accuracy rates for spontaneous speech over read speech (75% vs. 64.5%). Moftah & Roach found statistically similar identification rates for both laryngograph signal speech and low-pass filtered speech (64% vs. 65.5%).

Other researchers have used sinusoidal or pulse train re-synthesis to look at the adults' use of prosody in language discrimination (Atkinson, 1968; Ohala and Gilbert, 1979; Barkat et al., 1999). These re-synthesized sounds have the same pitch and amplitude contours as the original sounds. Atkinson (1968) tested discrimination of English-Spanish by American English speakers using an ABX discrimination task and an identification task. Ohala & Gilbert (1979) tested identification of English-Japanese-Cantonese, using native listeners from each language. Evidence of discrimination was observed in both studies. Barkat et al. (1999) looked at the discrimination of Eastern and Western dialects of Arabic by native and non-native listeners. They found that native listeners could discriminate Eastern and Western dialects of Arabic (accuracy rate of 58%), but non-native listeners could not (accuracy rate of 49%).

Another method that has been used to isolate prosodic cues from segmental cues makes use of Linear Predictive Coding (LPC). By applying inverse-LPC-filtering to a signal, all the spectral information can be removed, leaving only prosodic information remaining. Navrátil (2001) used inverse-LPC-filtered speech to test language identification using sentences from Chinese, English, French, German and Japanese. Participants in this study were from a variety of linguistic backgrounds. Overall, Navratil found identification rates of 49% (chance here was 20%). Komatsu et al. (2002) also used inverse-LPC-filtering, coupled with a low-pass filter, to test discrimination between English and Japanese by Japanese listeners and also found significant performance. Accuracy rates in this study were 70%.

In sum, low-pass filtering, laryngograph signals, sinusoidal or pulse train re-synthesis, inverse-LPC-filtering, *saltanaj* re-synthesis, and sentence scrambling are all different techniques used to eliminate segmental information from speech stimuli. These various studies indicate that both adults and infants are quite capable of discriminating between languages using only prosodic information. However, these techniques do not distinguish different aspects of prosody from each other. Such aspects include rhythm and intonation.

1.2.2 Prosodic cues for language discrimination – Rhythm

Languages have frequently been classified in terms of their rhythm, since Pike (1945) and Abercrombie (1967), as either "stress-timed" or "syllable-timed," or if a more continuous classification scheme is assumed, somewhere in between. For example, English, Dutch, German and Russian are considered to have similar rhythm and have been classified as stress-timed. French, Spanish, Italian and Catalan have been classified as syllable-timed and are considered to have a rhythm that is distinct from stress-timed languages. Japanese is frequently considered to be representative of a third class of mora-timed languages.

It has been suggested that membership in these classes affects the way a language is processed by its native speakers. For example, adults speaking French, Spanish, Catalan and Portuguese – all syllable-timed languages – seem to segment their language into syllables (Mehler et al., 1981; Morais et al., 1989; Sebastian-Galles et al., 1992;

Bradley et al., 1993; Pallier et al., 1993), and Japanese and Telugu speakers into morae (Otake et al., 1993; Cutler & Otake, 1994; Murty, Otake & Cutler, 2007). Speakers of stress-timed languages, like English and Dutch, on the other hand, segment their language into stress-based units of strong-weak syllables, or feet (Cutler et al., 1986; Cutler and Norris, 1988). In fact, research shows that speakers use their native language-based segmentation strategy even when listening to non-native languages (Cutler et al., 1986; Otake et al., 1993).

The idea that language rhythm influences processing strategies has shaped the research on infants' discrimination of languages. Nazzi et al. (1998) observed that young infants (newborns to 2-month-olds) only seem to discriminate between languages when they come from different rhythm classes. For example, infants have been shown to discriminate stress-timed and syllable-timed languages – e.g., Russian & French, English & Italian (Mehler et al., 1988), English & Spanish (Moon, Cooper & Fifer, 1993), and English & French (Dehaene-Lambertz & Houston, 1998). They have been shown to discriminate between stress-timed and mora-timed languages – e.g., English & Japanese (Christophe & Morton, 1998). Infants have also been shown to fail to discriminate two languages within the same rhythm class, e.g., two stress-timed languages, English & Dutch (Christophe & Morton, 1998; Ramus, 2002). This observation led Nazzi et al. to propose the *rhythm hypothesis*, which suggests that infants "extract rhythmic properties from speech and sort sentences into a small number of classes or sets based on rhythmic, timing properties" (Nazzi et al., 1998; see also Mehler et al., 1996, for a similar proposal).

To support this hypothesis, Nazzi et al. (1998) tested French-learning newborns on their ability to discriminate between low-pass filtered utterances from British English & Japanese, and British English & Dutch. Discrimination was observed for English & Japanese, which represent two different rhythm classes, but not for English & Dutch, which are both stress-timed languages. They also tested discrimination between different combinations of languages: English, Dutch, Spanish and Italian. Discrimination was only observed when a combination of English & Dutch, two stress-timed languages, was contrasted with a combination of Spanish & Italian, two syllable-timed languages. Infants failed to discriminate other combinations, such as English & Spanish vs. Dutch & Italian, where a stress-timed language was grouped with a syllable-timed language.

As language discrimination ability develops, the reliance on rhythmic information is thought to be maintained. Nazzi et al. (2000) showed that American English-learning 5-month-olds could still discriminate languages across rhythm classes. They found evidence of discrimination between a stress-timed and mora-timed language, British English & Japanese, and between a syllable-timed and a mora-timed language, Italian & Japanese. Because infants have learned about the properties of their native language by this age, they were also able to discriminate between British English & Dutch and between American & British English, all stress-timed languages. However, infants failed to discriminate two non-native, rhythmically-similar languages, Italian & Spanish, two syllable-timed languages, and Dutch & German, two stress-timed languages. To account for this, Nazzi et al. (2000) suggested a corollary to the rhythm hypothesis – that older infants have specifically homed into the rhythmic properties of their native language,

allowing them to discriminate it from other rhythmically similar languages. For non-native language pairs, 5-month-olds should only discriminate two languages from different rhythm classes.

1.2.3 Limitations of an account based on sensitivity to rhythm class

Initially, the idea of rhythm classes was based on the idea of isochrony – that stress-timed languages placed stressed syllables at regular intervals and syllable-timed languages did the same for all syllables. However, research seeking to prove this isochrony in production data has not been fruitful (see Beckman, 1992; and Kohler, 2009 for a review).

It was later suggested that differences in language rhythm arise as a consequence of the specific phonological properties of a language, such as the phonotactic permissiveness of consonant clusters, the presence or absence of contrastive vowel length, and vowel reduction (Dauer, 1983). A language that is stress-timed is likely to allow more complex consonant clusters, to have lengthened vowels in stressed syllables and reduced vowels in unstressed syllables. A syllable-timed language is more likely to restrict consonant clusters and show comparable vowel length over different syllables.

This line of thought has led to the development of a variety of rhythm metrics intended to categorize languages into classes using the durations of segmental intervals (Ramus et al., 1999; Grabe and Low, 2002; Wagner and Dellwo, 2004; White and Mattys, 2007). For example, Ramus et al. (1999) have suggested three simple interval

measures (IM) to distinguish between languages from different rhythm classes: %V, the proportion of vocalic segments within a sentence, and ΔC and ΔV, the standard deviation of the duration of consonantal and vocalic intervals, respectively. Grabe & Low (2002) have proposed the Pairwise Variability Index (PVI), which measures the variability in duration between successive vocalic and consonantal intervals.

Although Ramus et al.'s IMs are global, i.e., taken over a whole sentence or some larger unit, whereas the PVI only takes into account local variability between neighboring segments, the predictions using these two measures are largely similar. Canonical stress-timed languages have smaller vocalic percentages (%V) and greater variability in consonantal and vocalic durations (ΔC, ΔV, PVIc, PVIv), whereas canonical syllable-timed languages show the opposite pattern. Both IM and PVI correctly categorize English, German and Dutch as stress-timed and Spanish and French as syllable-timed. Since then, a number of researchers have equated linguistic rhythm with segmental duration and timing, the target of these rhythm metrics (Ramus and Mehler, 1999; Guasti, 2002; Ramus, 2002; Byers-Heinlein et al., 2010).

However, there have been several reported problems with these rhythm metrics. For example, IM and PVI measures occasionally produce differing predictions. IM measures categorize Luxembourgish as stress-timed and Thai as syllable-timed, whereas the reverse is obtained using PVI measures (Grabe & Low, 2002). In response to this, Ramus (2002a) suggested that a lack of control of speech rate was responsible for the different classifications. Speech rate can certainly affect the variability of segment durations, which can, in turn, affect the categorization by a rhythm metric. The efforts to

account for this have led to the proposal of the Varco measures (Dellwo, 2006). Varco, which divides the standard deviation of an interval measure (e.g., ΔC) by the mean duration of that interval, has been shown in some studies to successfully correct for rate of speech effects on IM measures (White and Mattys, 2007), but to eliminate language differences in others (Arvaniti, 2009).

Ramus (2002a) re-examined the materials from Ramus et al. (1999), which were controlled for speech rate, using the PVI measures, compared these results to the initial IM-based results, and found 'largely equivalent' results between the metrics. Even then, it is not clear which metrics should be used to classify languages. Among their measures, Ramus et al. (1999) suggest ΔC and %V provide the best classification. Using these dimensions, Ramus et al., grouped Polish with English and Dutch, other stress-timed languages. In contrast, with PVI, Polish is more closely grouped with the syllable-timed languages (Ramus, 2002a). In a study using large speech corpora, Loukina et al. (2010) tried to computationally identify languages using a large number of different rhythm measures and showed that no single metric or set of metrics is effective at classifying languages. Rather, they needed at least three metrics to identify the languages in their study, but showed that different measure triplets grouped languages differently. There was no consistent grouping based on stress-timed or syllable-timed.

Other researchers have found that these metrics are not robust to differences in preparation of source material or inter-speaker variability. Wiget et al. (2010) found that differences in sentence type caused significant variation in metric scores, for example, and Arvaniti (2009) found that classification by rhythm metrics can be pushed by cherry-

picking speech samples. Applying metrics to English sentences showing simple consonant-vowel alternations can cause it to be classified as syllable-timed. Additionally, even when applied to uncontrolled, spontaneous speech, inter-speaker variability in the metrics can prevent significant differences between languages, and thus, classification (Arvaniti, 2009). Indeed, based on their study using large corpora of speech, Loukina et al. (2010) suggest that, owing to inter-speaker variability, classifying languages reliably using just a single paragraph might be impossible.

Still other researchers have criticized these metrics as having moved away from the original notions of speech rhythm, and even of Dauer's analysis, on which they were originally based (Kohler, 2009), and have pointed out that they are more measures of speech timing than rhythm (Arvaniti, 2009). In fact, there may not be such a thing as neatly defined rhythm classes at all. Numerous languages have proved resistant to being neatly classified into these groups (Dauer, 1983; Ramus, 2002a; Arvaniti, 2009). Indeed, the notion that languages could be labeled exclusively as either stress-timed or syllable-timed originated after the introduction of those terms (Kohler, 2009). Languages were originally described as having both qualities, just with differing amounts (Pike, 1945). This idea has been resurrected again more recently by Nolan & Asu (2009), who showed that Estonian simultaneously has timing properties like Spanish, a syllable-timed language, and English, a stress-timed language.

It is evident from inspecting Table 1.3 that even in the developmental literature on language discrimination, there are examples that violate the rhythm hypothesis suggested

Infants' Native Language	Age	Tested Languages			Non-native Pair?	Evidence of Discrim.	Native Rhythm Class	Tested Languages Rhythm Class			Source
Canadian English	newborn	Canadian English	&	Tagalog		Yes	Stress	Stress	&	Syllable	Byers-Heinlein et al., 2010
Canadian Eng.-Tagalog biling.	newborn	Canadian English	&	Tagalog		Yes		Stress	&	Syllable	Byers-Heinlein et al., 2010
French	newborn	Italian	&	American English	Yes	Yes	Syllable	Syllable	&	Stress	Mehler & Christophe 1995
French	newborn	French	&	Russian		Yes	Syllable	Syllable	&	Stress	Mehler et al. 1988
Mixed	newborn	French	&	Russian	Yes	No		Syllable	&	Stress	Mehler et al. 1988
American English	newborn	American English	&	Spanish		Yes	Stress	Stress	&	Syllable	Moon, Cooper & Fifer 1993
Spanish	newborn	Spanish	&	American English		Yes	Syllable	Syllable	&	Stress	Moon, Cooper & Fifer 1993
French	newborn	Dutch	&	British English	Yes	No	Syllable	Stress	&	Stress	Nazzi, Bertoncini & Mehler 1998
French	newborn	Japanese	&	British English	Yes	Yes	Syllable	Mora	&	Stress	Nazzi, Bertoncini & Mehler 1998
French	newborn	Dutch	&	Japanese	Yes	No	Syllable	Stress	&	Mora	Ramus, 2002
French	newborn	Dutch	&	Japanese	Yes	Yes	Syllable	Stress	&	Mora	Ramus, 2002
British English	2-months	British English	&	Japanese		Yes	Stress	Stress	&	Mora	Christophe & Morton 1998
British English	2-months	Dutch	&	Japanese	Yes	No	Stress	Stress	&	Mora	Christophe & Morton 1998
British English	2-months	British English	&	Dutch		No	Stress	Stress	&	Stress	Christophe & Morton 1998
British English	2-months	French	&	Japanese	Yes	No	Stress	Syllable	&	Mora	Christophe & Morton 1998
American English	2-months	American English	&	French		Yes	Stress	Stress	&	Syllable	Dehaene-Lambertz & Houston 1998
French	2-months	American English	&	French		No	Syllable	Stress	&	Syllable	Dehaene-Lambertz & Houston 1998
American English	2-months	American English	&	Italian		Yes	Stress	Stress	&	Syllable	Mehler et al. 1988
American English	2-months	French	&	Russian	Yes	No	Stress	Syllable	&	Stress	Mehler et al. 1988
Catalan	4-months	Spanish	&	Catalan		Yes	Syllable	Syllable	&	Syllable	Bosch & Sebastian-Galles, 1997
Spanish	4-months	Spanish	&	Catalan		Yes	Syllable	Syllable	&	Syllable	Bosch & Sebastian-Galles, 1997
Catalan	4-months	Catalan	&	American English		Yes	Syllable	Syllable	&	Stress	Bosch & Sebastian-Galles, 1997
Spanish	4-months	Spanish	&	American English		Yes	Syllable	Syllable	&	Stress	Bosch & Sebastian-Galles, 1997
Spanish-Catalan bilingual	4-months	Maternal language	&	American English		Yes	Syllable	Syllable	&	Stress	Bosch & Sebastian-Galles, 1997
Spanish-Catalan bilingual	4-months	Maternal language	&	Italian		Yes	Syllable	Syllable	&	Syllable	Bosch & Sebastian-Galles, 1997
American English	5-months	American English	&	Spanish		Yes	Stress	Stress	&	Syllable	Bahrick and Pickens, 1988
American English	5-months	British English	&	Japanese		Yes	Stress	Stress	&	Mora	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	Italian	&	Japanese	Yes	Yes	Stress	Syllable	&	Mora	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	Spanish	&	Italian	Yes	No	Stress	Syllable	&	Syllable	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	British English	&	Dutch		Yes	Stress	Stress	&	Stress	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	German	&	Dutch	Yes	No	Stress	Stress	&	Stress	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	American English	&	British English	Yes	Yes	Stress	Stress	&	Stress	Nazzi, Jusczyk & Johnson, 2000
S.W. British English	5 & 7-months	S.W British English	&	Welsh English		Yes	Stress	Stress	&	Stress	Butler et al., 2010
S.W. British English	5-months	Scottish English	&	Welsh English	Yes	No	Stress	Stress	&	Stress	Butler et al., 2010
Australian English	3-16-months	Australian English	&	Japanese		Yes	Stress	Stress	&	Mora	Sundara et al. 2008
Australian English	3-16-months	Australian English	&	German		No	Stress	Stress	&	Stress	Sundara et al. 2008
Japanese	4-14-months	Japanese	&	?? English		Yes	Mora	Mora	&	Stress	Hayashi et al. 2001
French	newborn	British-Dutch	&	Spanish-Italian	Yes	Yes	Syllable	Stress-Stress	&	Syllable-Syllable	Nazzi, Bertoncini & Mehler 1998
French	newborn	British-Spanish	&	Dutch-Italian	Yes	No	Syllable	Stress-Syllable	&	Stress-Syllable	Nazzi, Bertoncini & Mehler 1998

Table 1.3. A summary of the language pairs that have been tested on infants in discrimination/preference studies, repeated from Table 2.1. The rhythmic classification for the native and tested languages is specified.

by Nazzi et al. (1998; 2000). For instance, in Dehaene-Lambertz & Houston (1998), French-learning 2-month-olds could not discriminate between English & French.[2]

In other studies, English-learning 2-month-olds have been reported to fail to discriminate between French and Russian, a supposed syllable-timed and stress-timed language, respectively (Mehler et al., 1988). They also fail to discriminate between French & Japanese, the latter of which has been classified as mora-timed, and even between Dutch & Japanese, a stress-timed and a mora-timed language (Christophe & Morton, 1998). Also, Mehler et al. (1988) found no evidence of discrimination between French and Russian by newborns without a French-learning background.

It has been suggested that these null results may be due to an insensitivity in the methodology used (Nazzi et al., 2000). Both Mehler et al. (1988) and Christophe & Morton (1998) used a between-subjects design. In the HAS habituation-change paradigm used in these experiments, infants were divided into two groups – for one group, the familiar language was presented following habituation, for the other, the novel language was presented. Nazzi et al. (2000) used a within-subjects design. In that study, each infant heard both the familiar and novel language during the test phase. In their study, Nazzi et al. (2000) reported positive evidence of discrimination for English & Japanese by American English-learning 5-month-olds. Yet, when they attempted the same experiment using a between-subjects habituation setup like Mehler et al. (1988) and Christophe & Morton (1998), they found no evidence of discrimination. Still, until the

[2] In that study, they suggest this may be due to a combination of a cross-linguistic preference for English and influence from the native langauge, French, leading to faster orientation times towards the native language. For French-learning infants, these factors cancel each other out.

language pairs can be tested using a within-subjects design, we cannot completely discount these null results.

Regardless of the controversy surrounding rhythm metrics and even the existence of rhythmic groups, the segment duration and timing properties that these metrics measure, and that some researchers, have equated with rhythm, are important characteristics of languages. Indeed, adult listeners have been shown to be sensitive to this information when discriminating languages, at least for rhythmically dissimilar languages like English & Japanese (Ramus & Mehler, 1999, discussed below). It is not clear, though, that listeners are sufficiently attuned to this type of information to discriminate between languages considered to be rhythmically similar.

1.2.4 Cues for language discrimination – Pitch

Intonational prominence is an additional aspect of speech prosody that is not captured by rhythm metrics, but likely plays a role in language discrimination. As with prosody, intonational prominence usually refers to a number of properties of language, both phonetic and phonological, like, for example, stress. Stress is a phonological property of language, associated with a specific syllable within a word (Lehiste, 1970). The location of stress within a word can change its meaning, like ['ɪm.pɔɹt], a noun, and [ɪm.'pɔɹt], a verb. When the word is placed in a sentence, the stressed syllable may or may not receive phonetic markings of various types, such as changes in amplitude, duration or spectral tilt. Stressed syllables are often marked with a specific type of pitch movement,

called a pitch accent. For example, English has stress and, in declarative sentences, typically marks the stressed syllable with a high pitch, or with a rising pitch (Dainora, 2001; Ananthakrishnan and Narayanan, 2008).

Not all languages have stress, but they all make use of pitch. In English, and other lexical stress languages, like German or Dutch, stressed syllables are optionally marked with pitch movements. In other languages, like Japanese, a specific syllable is always marked with a pitch movement whenever the word is produced; there is no optionality. Like in lexical stress languages, the placement of this pitch accent can change the meaning of the word. In Japanese, this movement is always a falling pitch contour (Beckman and Pierrehumbert, 1986; Pierrehumbert and Beckman, 1988). These languages are called lexical pitch accent languages. Other languages, like Mandarin, are tone languages. In tone languages, every syllable is marked with a specific pitch movement and the choice of contour affects the meaning of the word. Finally, some languages, like French, do not fall into any of these three classifications. In these languages, specific syllables are not phonologically marked to change the meaning of a word, yet words still carry a grammatically important pitch contour.

Languages use pitch to not only highlight specific syllables, but to mark the edges of phrases. These phrase-final or phrase-initial pitch contours are referred to as boundary tones. All known languages use pitch to mark the ends of large phrases (Jun, 2005). For example, English declarative sentences typically end in a low, falling pitch and questions typically end in a high, rising pitch. Yet, in some languages, like French and Japanese, pitch is used to mark the edge of word-sized phrases, as well. These languages are called

edge-marking languages. Languages in which a specific syllable within the word is marked with pitch, like English, are called head-marking languages. Languages can be categorized as edge marking, head marking or both. This typology is outlined in Jun (2005). Table 1.4 summarizes the intonational classification for all the language pairs used in the infant experiments reported in Table 1.1.

From Table 1.4 we can see that this intonational typology does not appear to correlate with the rhythm classification system of stress-, syllable- or mora-timed. For example, English and Spanish are both stress and head-marking languages, but English is considered stress-timed and Spanish is considered syllable-timed. Some researchers view intonational prominence as integral to the notion of linguistic rhythm (Kohler, 2009), but as stated above, many others have equated linguistic rhythm with segmental timing and duration alone (Ramus & Mehler, 1999; Ramus, 2002; Guasti, 2002). Some infant researchers explicitly discuss rhythm and pitch as separate phenomena (Nazzi and Ramus, 2003). For the remainder of this thesis, rhythm will refer to the segmental timing and duration patterns and properties, including speech rate and phrase-final lengthening and pauses, and intonation will refer explicitly to patterns and properties of pitch. Other prosodic properties, such as changes in amplitude, will not be significantly discussed.

It is possible that infants and adults are using pitch cues when they are discriminating between languages. Both infants and adults are certainly sensitive to pitch, even in non-tonal languages like English. Adults have to attend to pitch information in order to extract certain semantic and pragmatic information from speech,

Infants' Native Language	Age	Tested Languages			Non-native Pair?	Evidence of Discrim.	Native Lexical Intonation Type	Tested Languages Lexical Intonation Type			Native Inton. Marking Type	Tested Languages Lexical Intonation Type			Source
Canadian English	newborn	Canadian English	&	Tagalog		Yes	Stress	Stress	&	? Stress?	Head	Head	&	??	Byers-Heinlein et al., 2010
Canadian-Tagalog	newborn	Canadian English	&	Tagalog		Yes		Stress	&	? Stress?	Head	Head	&	??	Byers-Heinlein et al., 2010
French	newborn	Italian	&	American English	Yes	Yes	None	Stress	&	Stress	Head-Edge	Head	&	Head	Mehler & Christophe 1995
French	newborn	French	&	Russian		Yes	None	None	&	Stress	Head-Edge	Head, Edge	&	Head	Mehler et al. 1988
Mixed	newborn	French	&	Russian	Yes	No		None	&	Stress	Head-Edge	Head, Edge	&	Head	Mehler et al. 1988
American English	newborn	American English	&	Spanish		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Moon, Cooper & Fifer 1993
Spanish	newborn	Spanish	&	American English		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Moon, Cooper & Fifer 1993
French	newborn	Dutch	&	British English	Yes	No	None	Stress	&	Stress	Head-Edge	Head	&	Head	Nazzi, Bertoncini & Mehler 1998
French	newborn	Japanese	&	British English	Yes	Yes	None	Pitch Accent	&	Stress	Head-Edge	Head, Edge	&	Head	Nazzi, Bertoncini & Mehler 1998
French	newborn	Dutch	&	Japanese	Yes	No	None	Stress	&	Pitch Accent	Head-Edge	Head	&	Head, Edge	Ramus, 2002
French	newborn	Dutch	&	Japanese	Yes	Yes	None	Stress	&	Pitch Accent	Head-Edge	Head	&	Head, Edge	Ramus, 2002
British English	2-months	British English	&	Japanese		Yes	Stress	Stress	&	Pitch Accent	Head	Head	&	Head, Edge	Christophe & Morton 1998
British English	2-months	Dutch	&	Japanese	Yes	No	Stress	Stress	&	Pitch Accent	Head	Head	&	Head, Edge	Christophe & Morton 1998
British English	2-months	British English	&	Dutch		No	Stress	Stress	&	Stress	Head	Head	&	Head	Christophe & Morton 1998
British English	2-months	French	&	Japanese	Yes	No	Stress	None	&	Pitch Accent	Head	Head, Edge	&	Head, Edge	Christophe & Morton 1998
American English	2-months	American English	&	French		Yes	Stress	Stress	&	None	Head	Head	&	Head, Edge	Dehaene-Lambertz & Houston 1998
French	2-months	American English	&	French		No	None	Stress	&	None	Head-Edge	Head	&	Head, Edge	Dehaene-Lambertz & Houston 1998
American English	2-months	American English	&	Italian		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Mehler et al. 1988
American English	2-months	French	&	Russian	Yes	No	Stress	None	&	Stress	Head	Head, Edge	&	Head	Mehler et al. 1988
Catalan	4-months	Spanish	&	Catalan		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Bosch & Sebastian-Galles, 1997
Spanish	4-months	Spanish	&	Catalan		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Bosch & Sebastian-Galles, 1997
Catalan	4-months	Catalan	&	American English		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Bosch & Sebastian-Galles, 1997
Spanish	4-months	Spanish	&	American English		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Bosch & Sebastian-Galles, 1997
Spanish-Catalan	4-months	Maternal lang.	&	American English		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Bosch & Sebastian-Galles, 1997
Spanish-Catalan	4-months	Maternal lang	&	Italian		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Bosch & Sebastian-Galles, 1997
American English	5-months	American English	&	Spanish		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Bahrick and Pickens, 1988
American English	5-months	British English	&	Japanese		Yes	Stress	Stress	&	Pitch Accent	Head	Head	&	Head, Edge	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	Italian	&	Japanese	Yes	Yes	Stress	Stress	&	Pitch Accent	Head	Head	&	Head, Edge	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	Spanish	&	Italian	Yes	No	Stress	Stress	&	Stress	Head	Head	&	Head	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	British English	&	Dutch		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	German	&	Dutch	Yes	No	Stress	Stress	&	Stress	Head	Head	&	Head	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	American English	&	British English	Yes	Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Nazzi, Jusczyk & Johnson, 2000
S.W. British Eng.	5 & 7-mos	S.W.Brit. English	&	Welsh English		Yes	Stress	Stress	&	Stress	Head	Head	&	Head	Butler et al. 2010
S.W. British Eng	5-months	Scottish English	&	Welsh English		No	Stress	Stress	&	Stress	Head	Head	&	Head	Butler et al. 2010
Australian English	3-16-mos	Australian English	&	Japanese		Yes	Stress	Stress	&	Pitch Accent	Head	Head	&	Head, Edge	Sundara et al. 2008
Australian English	3-16-mos	Australian English	&	German		No	Stress	Stress	&	Stress	Head	Head	&	Head	Sundara et al. 2008
Japanese	4-14-mos	Japanese	&	?? English		Yes	Pitch Accent	Pitch Accent	&	Stress	Head-Edge	Head, Edge	&	Head	Hayashi et al. 2001
French	newborn	British-Dutch	&	Spanish-Italian	Yes	Yes	None	Stress	&	Stress	Head-Edge	Head	&	Head	Nazzi, Bertoncini & Mehler 1998
French	newborn	British-Spanish	&	Dutch-Italian	Yes	No	None	Stress	&	Stress	Head-Edge	Head	&	Head	Nazzi, Bertoncini & Mehler 1998

Table 1.4. A summary of the language pairs that have been tested on infants in discrimination/preference studies, repeated from Tables 2.1 and 3.1. The intonational classification for the native and tested languages is specified.

including which words contain new discourse items or have focus, or whether a sentence is a statement or a question. Research has also shown that adults use pitch information to segment words from running speech (Reitveld, 1980; Bagou *et al.*, 2002; Welby, 2003; Kim, 2004; Tyler *et al.*, 2006; Kim and Cho, 2009; Tyler and Cutler, 2009; Warner et al., 2010; Choi and Mazuka, to appear).

The developmental literature indicates that infants' ability to perceive pitch becomes adult-like relatively early. At 3-months, frequency resolution around 1,000 Hz is adult-like, and resolution at higher frequencies matures by 6-months (Schneider *et al.*, 1990; Spetner and Olsho, 1990) such that pitch perception is adult-like in 7- to 8-month-olds (Clarkson and Clifton, 1985; Montgomery and Clarkson, 1997). More specific to a linguistic context, French newborns are able to discriminate between lists of Japanese words containing segmentally identical tokens which differ only in their pitch contour (Nazzi *et al.*, 1998). American English-learning 2-3-month-olds have been shown to discriminate between rising and falling pitch contours on syllables (Karzon and Nicholas, 1989), and American English-learning infants between 5- to 11-months of age can discriminate between words which vary only in the pitch on the final syllable, even if the difference in f0 is as low as 5 or 10 Hz (Bull *et al.*, 1985).

Not only are infants sensitive to pitch changes, they use pitch in processing speech. Infants prefer infant-directed speech over adult-directed speech because of the differences in pitch patterns, not because of differences in amplitude or duration (Fernald and Kuhl, 1987). Six-month-olds prefer to listen to sentences with pauses coincident with clause boundaries, indicating a preference for continuous over discontinuous pitch

contours (Hirsh-Pasek *et al.*, 1987). In fact, 6-month-old American English-learning infants weight pitch over other cues that mark clause boundaries, such as pauses or preboundary lengthening, although they require at least one of these secondary cues along with pitch to successfully segment clausal units (Seidl, 2007). This is a pattern that develops between 4- and 6-months, as 4-month-olds show no preference for pitch cues in this task (Seidl and Cristià, 2008). This is the same age infants develop the ability to discriminate between certain language pairs that were not distinguishable at birth (Nazzi et al., 2000).

These studies indicate that both infants and adults are sensitive to and use pitch in processing language. Based on this research, it is likely that pitch information factors into the discrimination ability of infants and adults. In fact, the infants' growing attention to pitch found in the segmentation literature could also explain the observed change in language discrimination abilities that occurs around the same age – an alternative to the rhythmic hypothesis proposed by Nazzi et al. (2000). The following section reviews the handful of studies that have attempted to disentangle the importance of different aspects of prosody in language discrimination. (# nice)

1.3 Previous research that unpacks prosodic influences in discrimination tasks

A systematic attempt to examine the contribution of different prosodic cues in adults' language discrimination was made in two studies by using experimental stimuli that effectively isolate different components of the prosodic signal, namely rhythmic timing

and pitch information. Ramus & Mehler (1999) used re-synthesized speech to test whether French-speaking adults could discriminate between English and Japanese using either rhythmic timing or F0 information, or a combination of the two. They developed three conditions to test adults' reliance on prosodic cues. In two conditions, all consonants were replaced with /s/ and all vowels with /a/. This was done to remove all information about segments from the speech stimuli, but maintain rhythmic information as measured by the IM metric (Ramus et al., 1999). *Sasasa* speech contained both rhythmic timing and pitch cues, and *flat sasasa* speech, with a re-synthesized, flat F0 contour, contained rhythmic timing cues, but not pitch cues. For the third condition, *aaaa* speech, all segments were replaced with /a/. This condition contained only pitch, no rhythmic information. The pitch contour was extrapolated over voiceless segments. Ramus et al. (1999) found that French adults could discriminate English and Japanese only when rhythmic timing cues were available (i.e., in both the *sasasa* and *flat sasasa* conditions). Pitch alone was not enough for successful discrimination, nor did its addition improve discrimination performance over using just rhythmic timing cues. Discrimination was at 68% (A'-score: 0.72) for the *flat sasasa*/rhythmic timing condition and 65% (A'-score: 0.68) for the *sasasa*/rhythmic timing + pitch condition.

In contrast to Ramus & Mehler's findings, there are reports in the literature that adults can discriminate between languages based on F0 differences. In a pilot study, Komatsu et al. (2004) synthesized sets of stimuli, using white noise and pulse trains (an unfiltered F0 source), to contain different combinations of three cues: F0, intensity, and Harmonics-to-Noise Ratio (HNR), a basic measure of voice quality. All but one of their

stimulus conditions contained a re-synthesized amplitude curve matching the original stimuli, from which rhythmic information can be derived. The stimulus condition that had no rhythmic timing information contained only pitch information. They synthesized stimuli corresponding to four languages, English, Spanish, Mandarin and Japanese, which differ both rhythmically and intonationally. Rhythmically, English is considered a stress-timed language, Spanish a syllable-timed and Japanese a mora-timed language (Ramus et al., 1999). The classification of Mandarin is unclear, considered to be either stress-timed (Komatsu et al., 2004) or syllable-timed (Grabe & Low, 2002). Intonationally, English and Spanish are both a lexical stress language, Japanese is a lexical pitch accent language, and Mandarin is a tone language. English and Spanish are both head-marking languages; Japanese is both a head and edge-marking language; and Mandarin is usually considered to be neither head nor edge-marking (Jun, 2005). Averaged across languages, Japanese-listeners were able to discriminate in all conditions. Discrimination was around 62% when either the rhythmic timing information (the stimuli using various combinations of amplitude and HNR) or pitch alone was available. When both rhythmic timing and pitch information was present, discrimination was much better, 75% - 79%.

Other studies have shown that pitch-based discrimination is possible even for prosodically-similar languages like English and Dutch (Willems, 1982; de Pijper, 1983), or Quebec and European French (Ménard et al., 1999). Willems (1982) had English (from Utrecht) and Dutch listeners identify British English sentences recorded by a single speaker. These sentences were taken from Halliday (1970) and spoken with either English intonation, matching Halliday (1970), or with Dutch intonation, matching the

contours of Dutch paraphrases of the same sentences. Both groups could successfully identify the intonation patterns. English listeners had an accuracy of 60%, and Dutch listeners had an accuracy of 76%. De Pijper (1983) also used sentences from Halliday (1970), but re-synthesized the pitch of the original tokens in various ways. Two conditions had simplified versions of the original English intonation, and one condition had Dutch intonation, based on a theoretical model of Dutch intonation (Collier and 't Hart, 1981). He had British English-listeners judge the re-synthesized and original sentences on a 5-point scale of acceptability. Sentences with re-synthesized Dutch pitch were rated significantly lower than sentences with English pitch (either the original, or the two re-synthesized conditions – there was no difference between these latter three). Ménard et al. (1999) had Quebec French listeners identify sentences as either Quebec or European French. Sentences were recorded by five speakers from each dialect, varying in length from 8 to 87 syllables, and were low-pass filtered. A preliminary acoustic analysis found that the filtered sentences differed significantly only in pitch. Subjects were successfully able to discriminate the two dialects.

This discrepancy in language discrimination studies based on pitch can be reconciled if one considers both the languages being tested and the language background of the experimental subjects. Like with infants, whether or not adults in these studies are native speakers of one of the languages to be discriminated, matters. In Ramus & Mehler (1999), French listeners were discriminating between English and Japanese, two non-native languages. However, in the other reviewed studies, listeners had to discriminate between their native language or dialect and a non-native language/dialect. Indeed,

Ramus & Mehler (1999) mention in their paper that, although French speakers could not use pitch cues alone to discriminate between English and Japanese, native English speakers could (A-score: 0.61).

Language familiarity was found to drive identification scores in several other studies as well (Muthusamy et al., 1994; Barkat & Vasilescu, 2001; Nazzi, 1997; Bond et al., 1998; Stockmal et al., 2000; Stockmal & Bond, 2002). For example, native Arabic speakers were able to discriminate between Eastern and Western Arabic dialects using only prosodic cues, but non-Arabic speakers were not (Barkat, Ohala & Pellegrino, 1999). Similarly, Western Arabic speakers were able to more reliably identify Western Arabic dialects than Eastern Arabic dialects, while the opposite was true for Eastern Arabic speakers (Barkat & Vasilescu, 2001). Thus it seems that adults are capable of using both rhythmic timing and pitch cues to discriminate between languages, and can use pitch cues to discriminate even prosodically similar languages, provided one of the language is their native language.

Only one study has looked at the relative contribution of different prosodic cues in infant language discrimination. Ramus (2002) used the same re-synthesis techniques from Ramus & Mehler (1999) to look at discrimination of Dutch & Japanese by French newborns. They used natural stimuli, *saltanaj* stimuli, which retain rhythmic, pitch and phonotactic cues, *sasasa* stimuli with artificial intonation, and *saltanaj* stimuli with artificial intonation. Oddly, they found that French newborns could discriminate Dutch & Japanese using *saltanaj* speech, but not with natural speech. They suggest this is due to the fact that their natural stimuli contained sentences from multiple speakers, which

might have distracted their newborn participants. Most previous discrimination experiments with newborns have used either a single multi-lingual speaker or low-pass filtered speech from multiple speakers; see Table 1.1. The exception to this is Moon, Cooper & Fifer (1993), where infants heard two speakers. According to Ramus & Mehler (1999), with the re-synthesized stimuli, infants could concentrate on the language differences, rather than across-speaker differences.

For the two experiments using re-synthesized stimuli without pitch cues (the *sasasa* and *saltanaj* conditions with artificial intonation), Ramus (2002) found only weak evidence of discrimination. Specifically, for the fake-intonation-*saltanaj* condition, Ramus only found discrimination during the first minute after the language change, whereas in the full *saltanaj* condition, discrimination held for the first two minutes. For the fake-intonation-*sasasa* condition, he found a difference in sucking rates during habituation, but not as a result of the change trials. In all of Ramus's (2002) experiments, French newborns consistently sucked more for re-synthesized Dutch stimuli during habituation trials. He claimed this difference in sucking rate during the habituation phase as evidence for discrimination.

This latter observation is similar to a preference found by Mehler et al. (1988) by French newborns for their native language over Russian. In their HAS habituation-change task, Mehler et al. (1988) observed that infants who were habituated to French showed a higher sucking rate throughout the habituation phase than infants habituated to Russian. This same trend was spotted in a later experiment using low-pass filtered stimuli, but the difference was not significant in this case. However, a preference of this

nature has not been observed in other language discrimination experiments. The experiments in both Mehler et al. (1988) and Ramus (2002) use between-subjects designs. It is not clear that the preferences seen in the habituation phases are anything more than cross-group variation in sucking rate.

Based on his experiments, Ramus (2002) concluded that infants could discriminate using only rhythmic information, but he acknowledged that his interpretation is limited by the weak evidence for discrimination. Recall that this is arguably an easy case for rhythm-based discrimination – a stress-timed language and a mora-timed language, two rhythmically distinct languages. While these experiments may indeed provide evidence that French newborns can use rhythmic information to discriminate between Dutch and Japanese, the poorer discrimination in the *sasasa* and *saltanaj* conditions with artificial intonation seems to indicate a heavy reliance on pitch information. The removal of pitch information reduced French newborns' ability to discriminate Dutch and Japanese. Indeed, Ramus even suggests that, "when dissociated from intonation, rhythm loses part of its salience," highlighting the importance of pitch to discrimination.

While there is evidence that adults can use pitch information to discriminate between prosodically-similar languages, as long as they are familiar with one of the languages, there has only been research examining the use of rhythmic information in discriminating prosodically-dissimilar languages. The use of rhythm in discriminating prosodically-similar languages has not yet been directly tested. In contrast, no studies

have attempted to disentangle the use of different aspects of prosody in discriminating between prosodically-similar languages by infants, or even in a native/non-native pair.

1.4 Outline of the following research.

The aim of this thesis is to explore both infants' and adults' ability to discriminate prosodically-similar languages using only prosodic cues. Specifically, it seeks to show that pitch plays an important role in language discrimination, and that listeners rely more on pitch information than on rhythmic information.

The remainder of this dissertation is divided into four chapters. In Chapter 2, I describe a series of language identification experiments with American English-speaking adults testing their ability to discriminate their native language from a non-native language, German, as well as from a non-native dialect, Australian English. These languages are prosodically similar – they're rhythmically considered to all be stress-timed and they have similar intonational systems. First, using acoustic analysis, I show subtle rhythmic and intonational differences between both pairs. Then, using re-synthesized speech, I show that adults are able to discriminate between both language pairs using only prosodic cues, and that they weight intonation over rhythmic information.

American English-learning infants' discrimination ability is tested in chapters 3 and 4, using the headturn preference paradigm. In Chapter 3, I describe a series of experiments on discrimination of American English and German. I show that the ability

to discriminate between these two languages develops between 5- and 7-months of age. Like adults, 7-month-olds can discriminate between these two languages using only prosodic cues, as evidenced by their ability to discriminate between low-pass filtered speech stimuli. Using stimuli with re-synthesized intonational contours, evidence is provided suggesting that infants also weight intonation over other cues such as rhythm.

It should also be noted that both chapters 2 and 3 are written as stand-alone papers, intended to be submitted directly to journals. As such, they both have their own complete introductory and concluding sections, repeating much of the material provided in the current chapter and the final chapter of this dissertation.

Using the same testing methodology from the experiments in chapter 3, in chapter 4, I present data on infants' discrimination of a number of additional language pairs, including American & Australian English, American English & Dutch, British English & Dutch, and American English & Japanese. Infants failed to discriminate these pairs in all but two of the tested cases – five-month-olds were able to discriminate between British English & Dutch and American English & Japanese. However, the behavior observed in these two cases does not fit with previous reports on infant discrimination. The ability to discriminate British English & Dutch was retested using a different methodology, visual habituation, and more typical behavior was observed. The implications for infant testing methodologies are discussed. Finally, in chapter 5, I summarize the findings of the current study and offer concluding remarks.

2. Language discrimination by adults

2.1 Introduction

The experiments described in the present paper examine the role of prosodic cues in language and dialect discrimination. The ability to discriminate languages appears very early in life (Christophe & Morton, 1998; Dehaene-Lambertz & Houston, 1997; Nazzi, Jusczyk & Johnson, 2000), even as early as birth (Mehler et al., 1988; Moon et al., 1993; Nazzi et al., 1998). Given that infants do not tune into the language-specific properties of consonant and vowel categories until later in the first year of life (Werker et al., 1981; Kuhl et al., 1992; Polka & Werker, 1994), it appears that segmental information may not be necessary for infants to discriminate some language pairs. With the use of low-pass filtering, a process which degrades segmental information but preserves prosody, researchers have confirmed that early in acquisition, infants use prosodic information to distinguish languages (Bosch & Sebastian-Galles, 1997; Mehler et al., 1988; Nazzi et al., 1998). This early sensitivity to prosodic information for discriminating and identifying languages and dialects continues in adulthood (see Komatsu, 2007 for a review).

Prosody is a cover term referring to several properties of language, including its intonational system and linguistic rhythm. Languages have frequently been described in terms of their rhythm, since Pike (1946) and Abercrombie (1967), as either "stress-timed" or "syllable-timed," or if a more continuous classification scheme is assumed, somewhere in between. Initially, this classification was based on the idea of isochrony. However,

research seeking to prove this isochrony in production data has not been fruitful (see Beckman, 1992; and Kohler, 2009 for a review).

More recently, it has been suggested that language rhythm arises from phonological properties of a language, such as the phonotactic permissiveness of consonant clusters, the presence or absence of contrastive vowel length and vowel reduction (Dauer, 1983). This line of thought has led to the development of a variety of rhythm metrics intended to categorize languages into rhythmic classes using measurements made on the duration of segmental intervals (Ramus et al., 1999; Grabe & Low, 2002; Wagner & Dellwo, 2004; White & Mattys, 2007). Although these various metrics have been shown to successfully differentiate between prototypical languages from different rhythm classes on controlled speech materials, they are less successful with uncontrolled materials and on non-prototypical languages (Ramus, 2002; Arvaniti, 2009; Wiget et al., 2010). Despite claims that these rhythm metrics are more measures of speech timing than of rhythm (Arvaniti, 2009), adult listeners have been shown to be sensitive to the durational and timing differences captured by such metrics when discriminating languages (Ramus & Mehler, 1999). Throughout the rest of this paper, when we talk about rhythmic timing information, we are referring to the segmental durational information of the sort captured by these various rhythm metrics.

Stress and intonational prominence are two additional aspects of speech prosody that are not captured by the rhythm metrics, but might play a role in language discrimination. Stress, which can be marked acoustically through, for example, changes in amplitude, duration or spectral tilt, is used by most languages. However, *all* languages

seem to make some use of intonation, or pitch. Pitch is heavily connected with stress in stress languages. For example, English has stress and often marks the stressed syllable with a specific pitch contour, most commonly a high pitch (Dainora, 2001; Ananthakrishnan & Narayanan, 2008). Pitch contours over the whole sentence consist of interpolated pitch between stressed syllables and phrase-final boundary contours. Languages with weak or no lexical stress still use pitch in systematic ways, often by marking word edges, as in Korean or French (Jun, 2005; Jun & Fougeron, 2000), making it a universally important component of the speech signal. Listeners presumably attend not only to their language's specific cues for stress, which are typically confined to a specific syllable, but also to the patterns of pitch, which span whole utterances, and can make use of these cues when discriminating between languages.

Although the research on infants and adults highlights the importance of prosody and its use by human listeners in language identification and discrimination, it remains unclear which sources of prosodic information people use, and if multiple sources are used, how they are integrated with one another. In this paper, we report on a series of acoustic and perceptual experiments to address these questions when applied to the discrimination of American English and German as well as American and Australian English.

2.1.1 Differentiating the components of prosody in language discrimination by Adult listeners

In perception experiments with adult listeners, several experimental techniques have been used to effectively isolate prosody from segmental information, either by filtering away high frequency information or by replicating prosodic information in new, synthesized sound files. These include low-pass filtering (Atkinson, 1968; Mugitani et al., 2000), use of laryngograph signals (Maidment, 1976; 1983; Moftah & Roach, 1988), sinusoidal or pulse train re-synthesis (Ohala & Gilbert, 1979; Barkat et al., 1999), and LPC re-synthesis or Inverse-LPC-filtering (Foil, 1986; Navratil, 2001; Komatsu et al., 2002). However, these techniques do not distinguish different aspects of prosody from each other.

Two studies have made a systematic attempt to disentangle the contribution of different prosodic cues in predicting language identification/discrimination by adult listeners by using experimental stimuli that effectively isolate different components of the prosodic signal in their stimuli, namely rhythmic timing and pitch information. Ramus & Mehler (1999) used re-synthesized speech to test whether French-speaking adults could discriminate between English and Japanese using either rhythmic timing or F0 information, or a combination of the two. They developed three conditions to test adults' reliance on prosodic cues. In two conditions, all consonants were replaced with /s/ and all vowels with /a/. This was done to remove all information about segments from the speech stimuli. *Sasasa* speech contained both rhythmic timing and pitch cues, and *flat*

sasasa speech, with a re-synthesized, flat F0 contour, contained rhythmic timing cues, but not pitch cues. For the third condition, *aaaa* speech, all segments were replaced with /a/. This condition contained only pitch, no rhythmic information. The pitch contour was extrapolated over voiceless segments. They found that French adults could discriminate English and Japanese only when rhythmic timing cues were available (i.e., in both the *sasasa* and *flat sasasa* conditions). Pitch alone was not enough for successful discrimination, nor did its addition improve discrimination performance using just rhythmic timing cues. Discrimination was at 68% (A'-score: 0.72) for the *flat sasasa*/rhythmic timing condition and 65% (A'-score: 0.68) for the *sasasa*/rhythmic timing and pitch condition.

In contrast to Ramus & Mehler's findings, there are reports in the literature that adults can discriminate between languages based on F0 differences. In a pilot study, Komatsu et al. (2004) synthesized sets of stimuli, using pulse trains and white noise, to contain different combinations of three cues: F0, intensity, and Harmonics-to-Noise Ratio (HNR). All but one of their stimulus conditions contained periods of noise and silence, thus maintaining durational or timing information of the type captured by rhythm metrics. The stimulus condition that had no rhythmic timing information contained only pitch information. They synthesized stimuli corresponding to four languages, English, Spanish, Mandarin and Japanese, which differ both rhythmically and intonationally. Rhythmically, English is considered a stress-timed language, Spanish a syllable-timed and Japanese a mora-timed language (Ramus et al., 1999). The classification of Mandarin is unclear, considered to be either stress-timed (Komatsu et al., 2004) or

syllable-timed (Grabe & Low, 2002). Intonationally, English and Spanish are both stress (i.e., post-lexical pitch accent) languages, Japanese is a lexical pitch accent language, and Mandarin is a tone language (Jun, 2005b). Averaged across languages, discrimination was possible in all conditions. Discrimination was around 62% when either the rhythmic timing information (the stimuli using various combinations of intensity and HNR) or pitch alone was available. When both rhythmic timing and pitch information was present, discrimination was much better, between 75% - 79%. Other studies have shown pitch-based discrimination is possible even for prosodically-similar languages like English and Dutch (Willems, 1982; de Pijper, 1983). This discrepancy in language discrimination studies based on pitch can be reconciled if one considers both the languages being tested and the language background of the experimental subjects.

When both rhythmic timing and pitch information are provided, regardless of their native language adult listeners are able to discriminate two languages that are considered to be from two different rhythm classes, such as English and Spanish (Atkinson, 1968), English and French (Maidment, 1976; 1983), or English and Japanese (Ohala & Gilbert, 1979; Ramus & Mehler, 1999). As mentioned previously, English has been described as a stress-timed language, Spanish and French as syllable-timed languages, and Japanese as a mora-timed language. However, when languages are rhythmically similar, the language background of the test subjects comes into play. For example, native Arabic speakers were able to discriminate between Eastern and Western Arabic dialects using only prosodic cues, but non-Arabic speakers were not (Barkat, Ohala & Pellegrino, 1999).

Similarly, adults are able to discriminate languages that are intonationally very distinct, like English, a post-lexical pitch accent language, and Chinese, a lexical tone language (Ohala & Gilbert, 1979). Again, language background has an effect on listener's discrimination ability only when the test languages have similar intonation systems. French speakers can not use pitch cues alone to discriminate between English and Japanese, but native English speakers can (Ramus & Mehler, 1999). However, French speakers are able to use intonational differences to discriminate and identify Quebec and European French (Mernard et al., 1999)

Thus, it seems that listeners must be native speakers of one of the test languages when using prosodic information to discriminate between a pair of rhythmically or intonationally similar languages. This has been shown to be the case for 5-month-olds as well (Nazzi, Jusczyk & Johnson, 2000): 5-month-olds can discriminate two languages from the same rhythm class as long as one is their native language. In contrast, 5-month-olds fail to discriminate two rhythmically and intonationally similar non-native languages, whether they come from their native rhythm class or not.

2.1.2 Aims of the current study

In this study, we tested whether American English-speaking adults could discriminate their native language and a prosodically-similar foreign language, German, as well as a foreign dialect, Australian English, when segmental information is unavailable. Our goal was to determine what types of prosodic information were necessary for this task. Are

rhythmic timing alternations between segments, as captured by the various rhythm metrics, or just pitch information sufficient, or do listeners require additional cues to discriminate prosodically-similar languages?

English and German are historically closely related languages. They are rhythmically similar, both are considered stressed-timed languages, and they are intonationally similar, both have tonal phonologies with similar inventories, both tend to position stress word-initially, and both tend to mark stress with a high pitch. These similarities also hold for American English and Australian English, two dialects of the same language.

As stimuli for these experiments, we recorded several hundred sentences in American and Australian English and German, described below. In Section 2.2, we examine the stimuli acoustically to determine how American English differs from Australian English and from German in rhythm and intonation. In Section 2.3, we describe perception experiments designed to determine whether it is possible to discriminate between prosodically-similar languages/dialects using only prosodic cues, and which cues are necessary and sufficient for adult native English speakers to discriminate these language/dialect pairs.

2.2 Experiment 2.1: Acoustic-prosodic measures that distinguish between languages

To determine what types of prosodic information American English-speaking adults could potentially use to discriminate their native language from a prosodically-

similar foreign language, and from a foreign dialect of their native language, we analyzed acoustic differences between American and Australian English and German on two prosodic dimensions – rhythmic timing and pitch, using stepwise logistic regression.

2.2.1 Stimuli

39 English sentences from Nazzi et al. (1998) were recorded by eight female speakers of American English and eight female speakers of Australian English, then translated and recorded by eight female speakers of German in a sound-attenuated booth or quiet room at a sampling rate of 22050 Hz. Sentences had comparable number of syllables and overall durations, as shown in Table 2.1. Efforts were taken to minimize voice quality differences (pitch and timbre) both within and between dialects. 20 sentences from each speaker were selected to form the final stimulus set, with an effort to select for a lack of disfluencies and mispronunciations. These sentences formed a database of 160 sentences per language/dialect. Sentences in the database were also equalized for average intensity at 70 dB using Praat (Boersma & Weenik, 2006).

	Average number of syllables per sentence	Average sentence duration
American English	18 (2)	3.38 s (0.40)
Australian English	18 (2)	3.95 s (0.51)
German	18 (2)	3.66 s (0.63)

Table 2.1 – Average number of syllables per sentence and average sentence duration for the stimuli used in Experiment 2.1.

2.2.2 Acoustic measures

2.2.2.1 Rhythmic measures

As mentioned in Section 2.1, many rhythm metrics have been developed in an attempt to measure and quantify the rhythmic timing of languages (Ramus et al., 1999; Grabe & Low, 2002; Wagner & Dellwo, 2004; White & Mattys, 2007). All metrics have been shown to have strengths and weaknesses (Grabe & Low, 2002; Ramus, 2002; Arvaniti, 2009), and there has not been any perceptual research identifying which metric best represents what the listeners attend to. Rather than choose between them, we applied all available metrics to our data.

Rhythm metrics traditionally measure intervals of vowel and consonant segments. However, this division can be problematic, particularly in Germanic languages where sonorant consonants often serve as syllabic nuclei. For example, such a division labels the middle syllable in 'didn't hear' as part of a single consonantal interval, due to the fully syllabic /n/. Fortunately, the division into vowel and consonant intervals does not appear to be necessary for these metrics to be useful. When based on other divisions, such as voiced and unvoiced segments, rhythm metrics have still been shown to be successfully descriptive (Dellwo et al., 2007). For our data, we segmented and labeled intervals of sonorants and obstruents. As will become clear in the next section, we chose

this division primarily for the purposes of re-synthesis because sonorant segments are the segments that carry pitch information, while obstruents obscure pitch.[3]

We used eleven measures of rhythmic timing. For each sentence, we measured the mean percent sonorant interval duration (%S) and the standard deviation of both the obstruent intervals (ΔO) and sonorant intervals (ΔS), analogous to the measures from Ramus et al. (1999), as well as versions of the deviation values corrected for speech rate, VarcoS and VarcoO (Dellwo 2006; White & Mattys 2007). The Varco measures require the mean duration of both sonorant and obstruent intervals, which we also included in the analysis. We also measured the raw and normalized PVI values (rPVI and nPVI respectively) for both sonorant and obstruent intervals, analogous to Grabe & Low (2002).

2.2.2.2 Intonational Measures

Unlike rhythm metrics, there are no established metrics for qualifying intonational differences between languages. To operationalize intonation differences, for sonorant segments of each sentence, the only segments that carry pitch, we measured the minimum, maximum and mean pitch (see Baken & Orlikoff, 2000, for review), using

[3] To ensure this selection did not alter our results, we also segmented and measured our stimuli in terms of vowel/consonant intervals. This division did not result in a larger number of measures significantly discriminating either language/dialect pair. In fact, for American English and German, fewer rhythmic timing measures significantly differed between languages along a vowel/consonant division than a sonorant/obstruent division. In a logistic regression model, classification rates for the two language pairs were statistically similar for the two segmentation methods (American and German, $\chi^2(1) = 0.30$, p = n.s.; American and Australian English, $\chi^2(1) = 0.05$, p = n.s.). Thus, segmenting sonorant/obstruents rather than vowel/consonants does not significantly change classification based on the rhythm metrics

Praat. We also included the number of pitch rises in each sentence, the average rise height, and the average slope. We focused on pitch rises because all our sentences were declarative. In both dialects of English, and in German, stressed syllables in declarative sentences are typically marked with a high tone preceded by either a shallow rise or by a steep rise (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Grice et al., 2005).[4] By counting the number of rises, we expected to be able to capture differences between languages in the frequency of these pitch accents. Measures of the slope were expected to capture differences in pitch accent selection. A language that uses the shallow pitch rise frequently should have a lower average slope than languages which use a steeper rise more frequently.

2.2.3 Results

In Table 2.2, we present the means and standard deviations for each rhythm and intonation measure for American English, Australian English and German. Whether or not listeners are specifically using the information captured by the various rhythm or intonation metrics, there are significant differences between the two language pairs in both rhythmic timing and pitch measures as compared using t-tests.

To test these differences further, we conducted a stepwise, binary logistic regression for each language pair in order to see how much of the data could be correctly

[4] In the ToBI-transcription system, a shallow rise to a high tone would be labeled as a H*, and a steep rise would be labeled as a L+H*, or occasionally a L*+H.

	American English	Australian English	American vs. Australian English	German	American English vs. German
%Son	59.51% (5.61)	58.72% (5.91)	t(318) = 1.229, p=n.s.	54.72% (6.76)	t(318) = 6.900, p<0.001
sd Son	93.16 (30.04)	110.19 (36.02)	t(318) = 4.592, p<0.001	77.73 (35.37)	t(318) = 4.208, p<0.001
sd Obs	49.58 (15.82)	59.98 (18.92)	t(318) = 5.334, p<0.001	58.07 (16.57)	t(318) = 4.690, p<0.001
rPVI Obs	57.46 (17.34)	66.73 (23.39)	t(318) = 4.027, p<0.001	66.37 (19.11)	t(318) = 4.370, p<0.001
nPVI Obs	65.35 (15.70)	59.68 (15.82)	t(318) = 3.219, p=0.001	66.25 (15.53)	t(318) = 0.514, p=n.s.
Mean Obs	93.43 (14.72)	113.23 (18.55)	t(318) = 10.577, p<0.001	102.74 (18.65)	t(318) = 4.955, p<0.001
rPVI Son	104.31 (37.59)	124.72 (45.56)	t(318) = 4.369, p<0.001	85.03 (39.56)	t(318) = 4.471, p<0.001
nPVI Son	71.74 (16.43)	73.55 (17.19)	t(318) = 0.961, p=n.s.	62.15 (13.68)	t(318) = 5.674, p<0.001
Mean Son	141.17 (28.94)	165.87 (37.26)	t(318) = 6.624, p<0.001	128.65 (40.02)	t(318) = 3.206, p=0.001
Varco Obs	52.90 (12.71)	52.74 (13.20)	t(318) = 0.106, p=n.s.	56.81 (12.51)	t(318) = 2.777, p<0.006
Varco Son	65.52 (13.50)	65.96 (12.32)	t(318) = 0.308, p=n.s.	59.10 (12.51)	t(318) = 4.415, p<0.001
Min F0	117.35 (39.95)	126.97 (48.35)	t(318) = 1.941, p=n.s.	114.6 (29.26)	t(318) = 0.701, p=n.s.
Max F0	320.35 (46.33)	303.12 (51.72)	t(318) = 30138, p=0.002	358.9 (73.30)	t(318) = 5.629, p<0.001
Mean F0	211.97 (18.72)	208.62 (29.41)	t(318) = 1.216, p=n.s.	195.03 (19.39)	t(318) = 7.949, p<0.001
Number of Rises	7.52 (2.47)	10.55 (2.70)	t(318) = 10.498, p<0.001	9.18 (2.64)	t(318) = 5.828, p<0.001
Average Rise (F0)	39.41 (12.62)	36.24 (11.14)	t(318) = 2.382, p=0.018	55.35 (23.02)	t(318) = 7.682, p<0.001
Average Slope	506.5 (493.30)	491.49 (434.80)	t(318) = 0.288, p=n.s.	1137.24 (1384.41)	t(318) = 5.429, p<0.001

Table 2.2. Means and standard deviations for each rhythm and pitch measure for the sentence stimulus set in American English, Australian English and German. T-test comparisons between American and Australian English and American English and German for each measure are also presented.

classified using these measures. American English was separately compared to German and Australian. We used logistic regression as an alternative to discriminant analysis because it requires fewer assumptions. Namely, logistic regression does not require independent variables to be normally distributed or have equal within-group variances. First, the 11 rhythm measures described above were used as independent variables. Classification scores are reported in Table 2.3. Using rhythm measures alone, the model was able to accurately classify the two pairs around 73% of the time. This is well above chance, and somewhat surprising, considering the three tested languages are all stressed timed, and so expected to be rhythmically very similar.

	Rhythmic Timing	Pitch	Rhythm & Pitch
American vs. Australian English	76.6%	78.8%	87.8%
American English vs. German	70.2%	86.3%	89.2%

TABLE 2.3. Classification scores from a logistic regression for the two language/dialect pairs under different conditions: rhythmic timing information only, pitch information only, and the combination of rhythmic timing and pitch information.

We also ran logistic regressions testing the classification using only pitch measures. These results are also presented in Table 2.3. For American English vs. German, classification was higher using pitch measures when compared to rhythmic timing measures ($\chi^2(1) = 7.61$, $p = 0.006$). However, for American vs. Australian English, classification using only pitch was comparable to classification using only rhythmic timing ($\chi^2(1) = 0.35$, $p = 0.554$.). Although the intonational systems of the three languages are similar in their tonal inventories, differences were expected in how these

tones were employed by the different languages. We take the high classification as support for the existence of such differences.

Finally, although classification using just rhythmic timing measures or just pitch measures was high, it was expected that when the regression model had access to both types of measures, it would perform even better. Surprisingly, classification was not significantly higher for either language pair when rhythm measures were supplied to the model in addition to the pitch information (American English vs. German, $\chi^2(1) = 0.35$, $p = 0.554$; American vs. Australian English, $\chi^2(1) = 2.91$, $p = 0.088$). However, access to both cues did allow for greater classification accuracy than access to just rhythmic timing cues, (American English vs. German, $\chi^2(1) = 11.16$, $p < 0.001$; American vs. Australian English, $\chi^2(1) = 4.29$, $p = 0.038$).

In summary, logistic regression using acoustic measures of rhythmic timing and pitch showed that both language pairs could be classified using either cue type. Based on these acoustic differences, adult listeners should be able to discriminate between either American English and German or American and Australian English using rhythmic timing or pitch cues. However, based on the accuracy rates of our regression model, pitch cues should be more informative than rhythmic cues for both pairs.

2.3 Experiment 2.2 – The perceptual role of rhythmic timing and pitch

The previous section showed that the two language pairs, American vs. Australian English and American English vs. German, can be distinguished acoustically using only

rhythmic timing or pitch information. This section explores whether adult listeners can use prosodic information alone to distinguish American English vs. German and American vs. Australian English, and if so, which cues they use. To do this, we conducted four perceptual experiments, each testing a different combination of prosodic cues.

Traditionally, segmental information is removed from natural speech using low-pass filtering. Discrimination is tested on low-pass filtered speech (filtered above 400 Hz) in the first, *filtered,* condition. It is worth noting the objections made by Ramus & Mehler (1999) against low-pass filtering. On one hand, filtering does not actually remove all segmental information. While most segmental information is contained in frequencies above 400 Hz, some, such as the first formant of high vowels, can be located below that frequency. On the other hand, filtering above 400 Hz may remove some pitch information, especially for female speakers, like those used in this study. Our stimuli do contain tokens where the pitch ranges higher than 400 Hz. Thus, low-pass filtering may have distorted the acoustic features analyzed in Experiment 2.1.

In the second experimental condition, we tested discrimination using re-synthesized speech. Unlike low-pass filtering, re-synthesis completely removes segmental information and leaves rhythmic timing and pitch information intact. It also strips out other prosodic cues that low-pass filtering does not, like changes in amplitude and voice quality, making it a more controlled test case. The only cues available to

listeners in this condition are pitch contour information and rhythmic timing of sonorant and obstruent segments.[5] We refer to this condition as the *rhythm-and-pitch* condition.

In the third, *rhythmic timing only,* condition, we tested discrimination when only rhythmic timing information is available. For this condition, the pitch contours from the re-synthesized stimuli were removed, leaving a flat, monotone pitch.

Finally, in the fourth, *scrambled,* condition, we tested discrimination when neither rhythmic timing nor pitch information were available in their original, natural order. The re-synthesized stimuli from the second condition, which contain both rhythmic timing and pitch information, were divided into five portions and their temporal order was scrambled. Scrambling distorts the pitch contours and rhythmic patterns by destroying their original sequencing, and thus should remove any recognizable prosodic patterns.

Based on the acoustic analyses presented in the first experiment, we expected listeners to be able to discriminate both language pairs in the *filtered* condition, and the *rhythm-and-pitch* condition. Acoustically, there is enough information contained in the pitch contours and rhythmic timing patterns of the experimental stimuli to successfully discriminate the languages, and our acoustic analysis does not even take into account the

---

[5] To ensure that listeners were not significantly affected by the choice to re-synthesize around sonorant/obstruent segments, we also tested discrimination on monotone, re-synthesized sentences of American English and German, when the stimuli were re-synthesized along a vowel/consonant division. Acoustically, the sonorant/obstruent segmentation was slightly more informative than a vowel/consonant segmentation. As a result, we expected either no difference or a poorer performance in this condition than in the sonorant/obstruent re-synthesis condition. Indeed, group discrimination between American English and German using stimuli re-synthesized along a vowel/consonant division was poorer than stimuli using sonorant/obstruent segments (%Correct: 0.51; A': 0.51; comparison to chance: $t(14) = 0.4$; $p = 0.7$), but the number of individual subjects performing above chance was comparable (8/15; $\chi^2(1) = 1.29$, $p = 0.256$). These results indicate that discrimination using vowel/consonant re-synthesized stimuli is similar to, or slightly worse than discrimination using sonorant/obstruent re-synthesized stimuli. Thus, the decision to apply the rhythm metrics and re-synthesize along a sonorant/obstruent division at best overestimates the degree to which adults rely on rhythm for language discrimination.

other prosodic cues that may be present in the low-pass filtered stimuli. Subsequently, we tested subjects on the *rhythm timing only* condition to determine whether rhythmic timing cues are sufficient for successful discrimination. By the process of elimination, a failure to discriminate in the *rhythm timing only* condition would present evidence that listeners are attending to pitch information.

Finally, in the *scrambled* condition, we tested the hypothesis that prosodically-similar languages are better discriminated when listeners are native speakers of one of the discriminated languages. Native speakers presumably know and attend not only to the average durations of various segments or pitch contours associated with their language, but to the sequencing of this information as well. We expected that when this sequencing is disrupted in the *scrambled* condition, listeners will fail to discriminate American English vs. German as well as American vs. Australian English.

2.3.1 Methods

2.3.1.1 Stimuli

All 480 sentences from Experiment 2.1 were used for each condition (160 for each of the three languages). In the *filtered* condition, sentences were low-pass filtered using Praat at a frequency cut-off of 400 Hz, with 50 Hz smoothing.

In the *rhythm-and-pitch* condition, the sentences were re-synthesized. Sonorant segments were replaced with /a/ and obstruent segments were replaced with silence,

simulating a glottal stop, producing new sound files of the same length as the original. The method of re-synthesis was simpler to set up, and avoids any issues with co-articulation, or its absence, between the sonorant and obstruent segment because glottal stop and /a/ show no formant transitions. The original set of sentences had no disfluencies or sentence-medial pauses; therefore, all periods of silence in the re-synthesized stimuli corresponded to obstruent segments. However, there was no differentiation between sentence onset and offset consonants, and surrounding silence. Thus, information about any obstruents located on the sentence edges was lost in the re-synthesis. A new logistic regression model showed that classification scores for the languages were similar to those found in Section 2.2 (76.5% for American English and German, which is, surprisingly, a slight but non-significant improvement on the acoustic analysis previously reported in Section 2.2 (originally, 70.2%; χ^2 (1) = 1.02, p = 0.313), and 76.6% for American and Australian English, identical to the ones reported in Section 2.2. Nor did this change alter which rhythmic measures, shown in Table 2.2, the language pairs significantly differed on. Lastly, the resulting intonational contour from the sonorant segments in the original sentence was copied to the new file. The resulting re-synthesized stimuli thus contained both rhythmic timing and pitch information.

In the *rhythmic timing-only* condition, the re-synthesized sentences were modified in Praat to remove pitch. Sentences were given a flat, monotone pitch contour of 200 Hz. The resulting sentences, thus, only contained rhythmic timing information.

In the final, *scrambled*, condition, the re-synthesized sentences were scrambled in order to make rhythmic timing and pitch information uninformative, or at least,

unfamiliar. A Praat script was written to divide each sentence into five portions, which were then rearranged in a random order. Cuts between divisions were made at the onset of an obstruent interval (or the offset of a sonorant interval). To ensure that the scrambling only disrupted the original prosodic sequencing, but did not destroy any prosodic information that might still distinguish the languages, the rhythm and intonation of the resulting stimuli were re-measured as described in Experiment 2.1. Classification rates were unchanged with respect to the unmodified stimuli – for scrambled American English and German, 88.6% (originally 89.2%; $\chi^2(1) = 0.02$, $p = 0.888$), and for scrambled American and Australian English, 79.7% of the stimuli were correctly classified (originally, 87.8%; $\chi^2(1) = 2.41$, $p = 0.121$).

2.3.1.2 Participants

120 native speakers of American English were recruited from the undergraduate population of a university. Most received course extra credit, some were paid.

2.3.1.3 Procedure

30 subjects were used in each of the conditions, and for each condition, subjects were divided into two groups. 15 subjects heard American English vs. German and 15 heard American English vs. Australian English.

The experiments were presented in Praat in a sound-attenuated booth. Sentences were presented one at a time over loudspeakers at an average intensity of 70 dB. Each participant heard 320 sentences, 160 in each of the two languages/dialects. Sentences were presented in a randomized order. Subjects were told they would hear a number of sentences and had to decide if they were spoken in American English or some other language/dialect. They were not informed about either the number of foreign languages, dialects, or their identity. After each sentence was played, participants identified it as "American English" or "Other." Testing lasted for around thirty minutes.

2.3.2 Results

Percent correct scores for each condition are presented in Table 2.4. To take into account any response bias subjects may have had, we converted the responses into hit-rates and false alarm-rates. Correctly identified American English sentences were counted as hits, sentences misidentified as American English were counted as false alarms. Discrimination scores (A') were then calculated, and are also presented in Table 2.4. A'-scores are a non-parametric analog of d'-scores. They range from 0 to 1, where chance performance is 0.5. Analysis of percent correct data and A' show identical patterns, thus throughout the paper, only analyses with A' are reported. A one-sample t-test was conducted to compare the group A' scores to chance (0.5).

Subjects' data were also examined individually to see if they scored above

Experiment Condition	Language / dialect Pair	Percent Correct	A' score	Participants Above Chance	Comparison to Chance
Filtered	vs. Australian	0.55	0.58	11/15	$t(14) = 4.1, p = .001$
	vs. German	0.53	0.56	8/15	$t(14) = 3.0, p = .011$
Rhythm and Pitch	vs. Australian	0.55	0.59	13/15	$t(14) = 3.2, p = .007$
	vs. German	0.50	0.49	3/15	$t(14) = 0.5$, n.s.
Rhythm Only	vs. Australian	0.53	0.55	6/15	$t(14) = 2.0$, n.s.
	vs. German	0.56	0.59	11/15	$t(14) = 3.6, p = .003$
Scrambled	vs. Australian	0.51	0.52	5/15	$t(14) = 0.9$, n.s.
	vs. German	0.48	0.43	3/15	$t(14) = 2.0$, n.s.

TABLE 2.4. Results from Experiment 2.2, reporting group percent correct scores, A'-scores, the number of participants performing above chance and statistical tests comparing A'-scores to chance (0.5). The *filtered* condition tests discrimination using low-pass filtered sentences of American and Australian English, and American English and German. The *rhythm-and-pitch* condition tests discrimination using re-synthesized sentences containing both rhythmic timing and intonational cues. The *rhythm-only* condition tests discrimination using re-synthesized sentences containing only rhythmic timing information. The *scrambled* condition tests discrimination using sentences containing disrupted rhythmic timing and pitch information.

chance, in order to determine that individual performance conforms to group trends. To determine whether a subject performed significantly above chance, the 95% confidence limits were calculated based on the normal approximation to the binomial distribution (Boothroyd, 1984; Kishon-Rabin, Haras, & Bergman, 1997).[6] The confidence limits were calculated based on the number of trials (n = 320), the probability of guessing (1/2) and the t-value (2, if number of trials is more than 20). These parameters hold for all following experiments. Subjects with A'-scores above 0.556 were considered to perform significantly above chance with a $p < 0.05$. These results are also presented in Table III. Finally, the number of subjects performing above chance was compared across conditions using a chi-square test.

2.3.3 Discussion

For the *filtered* and *rhythm-and-pitch* conditions, we expected subjects to successfully discriminate the languages despite the rhythmic and intonational similarity of American English, Australian English and German, as all subjects were native speakers of

[6] Confidence limits for scores expected from guessing depend on the number of trials (n), and on the probability of guessing (p; in a two-alternative forced-choice, p = ½):

$$\text{Confidence Limit} = \text{Chance score} \pm t * \sqrt{\frac{p(1-p)}{n}}$$

Values for t are taken from the t-tables according to n – 1 degrees of freedom. For n = 20 or more, t = 2.0 for 95% confidence limits and 3.3 for 99.9% confidence limits. In our experiments, chance is at 0.5, thus:

$$95\% \text{ confidence limit} = 0.5 \pm \sqrt{\frac{1}{n}} \qquad 99.9\% \text{ confidence limit} = 0.5 \pm \frac{3.3}{2} * \sqrt{\frac{1}{n}}$$

American English, and because both language pairs were distinguishable acoustically, as shown in Experiment 2.1.

American English-speaking adults were able to discriminate their native language from a foreign dialect, Australian English, in both the low-pass *filtered* and *rhythm-and-pitch* conditions. There was no difference in performance between the two conditions (A' 0.58 and 0.59, respectively; $\chi^2(1) = 0.83$, $p = 0.362$). Thus, for American and Australian English, the rhythmic timing and pitch information present in both the re-synthesized and low-pass filtered stimuli is sufficient to allow for discrimination between the two dialects.

In both conditions, the task was quite difficult, indicated by the proximity of the average accuracy to chance. There are two possible explanations for this fact. First, it is possible that listeners do not maximally use the information present in the stimuli. They simply may not be capable of noticing the fine prosodic differences between the languages captured by the acoustic analysis, but rather rely more on segmental information to discriminate in more natural tasks. Second, the measures used in the acoustic analysis may not accurately capture the information perceptually extracted by human listeners.

Although as a group, listeners were able to discriminate American and Australian English in both the *filtered* and *rhythm-and-pitch* conditions, they failed to discriminate between the two dialects in the *rhythmic timing-only* condition. Consistent with the group results, a significantly greater number of subjects performed above chance in the *rhythm-and-pitch* condition compared to the *rhythm-only* condition ($\chi^2(1) = 7.03$, $p =$

0.008), and the difference between the *filtered* condition and the *rhythm-only* condition showed a trend towards significance ($\chi^2 = 3.39$, $p = 0.066$). These results indicate that pitch information is necessary for American English listeners to discriminate between their native dialect and Australian English; rhythmic timing information alone is not sufficient.

However, results for American English and German did not follow this pattern. Discrimination between the two languages was significantly above chance in the *filtered* condition, but not in the *rhythm-and-pitch* condition. The difference in the number of subjects performing above chance between the two conditions trended towards significance ($\chi^2 = 3.59$, $p = 0.058$). By themselves, these results might suggest that the rhythmic timing and pitch information in the re-synthesized stimuli are not sufficient to allow discrimination, but rather, unlike for American and Australian English, subjects required the additional cues still present in the filtered stimuli, such as amplitude or voice quality, to discriminate between American English and German.

Surprisingly, though, discrimination between American English and German was significantly above chance in the *rhythmic timing-only* condition, and significantly more listeners performed above chance in this condition than in the *rhythm-and-pitch* condition ($\chi^2(1) = 8.57$, $p = 0.003$). This indicates that the rhythmic timing information present in both re-synthesis conditions was sufficient to allow for discrimination, but the additional presence of pitch information somehow precluded discrimination of the two languages in the *rhythm-and-pitch* condition. This is surprising because, based on our acoustic measures, American English and German sentences were better classified using pitch

information alone, or using pitch and rhythmic timing information together, than using rhythmic timing information by itself. Yet, perceptually, the presence of pitch hindered listeners' discrimination ability, rather than improved it, contrary to the results of the acoustic study. Thus, it is likely that the pitch measures we developed did not capture all the properties of the speech signal listeners attend to. For example, our acoustic measures fail to capture any information about sequential ordering of pitch movements.

The inability to discriminate between American English and German in the *rhythm-and-pitch* condition, is likely due to an inability to discriminate between the pitch contours of these two languages. This may be because listeners view them as too similar to one another. Furthermore, although the languages can be discriminated using differences in rhythmic timing, the presence of indistinguishable pitch cues prevents listeners from effectively making use of the rhythmic cues. Thus, although subjects may be capable of using both cue types to discriminate American English and German, pitch information is weighted above rhythmic timing information.

Finally, for the *scrambled* condition, we expected subjects to fail to discriminate both language pairs. Previous research suggests that discrimination between prosodically-similar languages requires subjects to be a native speaker of at least one of the test languages (Barkat, Ohala & Pellegrino, 1999; Ramus & Mehler, 1999). Scrambling the stimuli should destroy any native pitch or rhythmic timing patterns American English listeners might be attending to by altering the sequencing of this information. Although, acoustically, the scrambled languages are still distinguishable, none of them should sound like a familiar language.

Indeed, subjects were unable to discriminate either language pair, and the number of subjects performing above chance was significantly lower than other re-synthesis conditions where successful discrimination was observed (for American vs. Australian, $\chi^2(1) = 8.89$, $p = 0.003$; for American vs. German, $\chi^2(1) = 8.57$, $p = 0.003$). Altering the sequencing of natural rhythmic timing and pitch information of the two languages, rendered them unfamiliar and thus, prevented listeners from discriminating between them.

2.4 General discussion

In this study, we examined whether it was possible to discriminate between closely related languages – American English and German, and American and Australian English – using prosodic cues alone. Using a logistic regression analysis, we showed that the two language pairs were acoustically distinct, in both rhythmic timing and pitch. Classification accuracy was lowest with rhythmic timing cues alone, whereas the model performed best when rhythmic timing and pitch were both included or with pitch-alone.

Next, with perception experiments involving low-pass filtered and re-synthesized stimuli, we demonstrated that American English listeners could discriminate between both language/dialect pairs using prosodic cues alone, though with difficulty. When most (*filtered* condition) or all (*rhythm-and-pitch* condition) of the segmental information had been stripped from the stimuli, successful discrimination was only slightly better than chance levels. When above chance, listeners in our experiments scored around 55%

correct (A'-score of around 0.58-0.6) for both language pairs. Our results are comparable to previous discrimination studies on prosodically-similar languages. Using Eastern and Western Arabic stimuli that had been re-synthesized using sinusoidal pulses simulating pitch, amplitude and rhythmic timing, Barkat, Ohala & Pellegrino (1999) found that native Arabic listeners correctly identified the dialects 58% of the time.

Crucially, in this study, we examined which prosodic cue – rhythmic timing or pitch – was used in discrimination. For American and Australian English, adult listeners were able to discriminate re-synthesized stimuli containing both rhythmic timing and pitch information, but when pitch was removed, they failed. Pitch is thus a necessary prosodic cue for discriminating between American and Australian English; rhythmic timing alone is not sufficient.

A different pattern was observed for American English and German. Adult listeners did not discriminate these two languages when pitch and rhythmic timing cues were available. Yet, when all pitch cues were removed, adult listeners successfully discriminated American English and German. Thus, for English and German, rhythmic timing alone is sufficient to cue discrimination; however, the presence of pitch cues interferes with discrimination.

Several conclusions can be drawn from these results. First, the pitch measures developed in this study do not capture all the information used by listeners to discriminate languages. Specifically, sequencing information appears to be important, since the distortion of this information can prevent discrimination. Second, perhaps surprisingly, the intonation systems of American English and German are too similar to

be distinguished, unlike the intonation systems of American and Australian English. Lastly, and more importantly, American English listeners weigh intonational cues over rhythmic cues when discriminating prosodically-similar languages. This would explain why they failed to discriminate American English and German in the *rhythm-and-pitch* condition. If intonation is too similar, discrimination fails, even if the language pair is rhythmically distinguishable.

At first glance, our results are inconsistent with Ramus & Mehler's findings. Ramus & Mehler (1999) found that native French listeners could use rhythm, but not pitch, to discriminate English from Japanese. Pitch neither aided nor hurt discrimination over and above rhythm alone. However, recall that their study differs from the current study in that they tested discrimination of two non-native languages. Unlike English and German, English and Japanese are prosodically quite distinct. English is considered a prototypical stress-timed language, and Japanese is considered a prototypical mora-timed language. Thus, Ramus & Mehler's results show that adult listeners can use a large difference in rhythm to discriminate language pairs, even when only the segment durational and rhythmic timing differences are present. Discrimination in this case is possible even when the pitch contours of the languages are unfamiliar or confusing.

Infant and animal research also supports the idea that large differences in rhythmic timing can be discriminated regardless of language familiarity. Several studies have shown that rhythmically distinct language pairs can be discriminated by newborns and young infants, regardless of infants' language background (Mehler et al., 1988; Dehaene-Lambertz & Houston, 1998; Moon et al., 1993, Mehler & Christophe, 1995;

Nazzi, Bertoncini, & Mehler, 1998). We also know that newborns share the ability to discriminate languages from different rhythm classes with cotton-top Tamarin monkeys (Ramus, Hauser, Miller, Morris & Mehler, 2000), as well as rats (Toro, Trobalón & Sebastián-Gallés, 2003). Thus, the ability to discriminate languages from different rhythm classes appears to be based on a species-general auditory sensitivity to rhythmic timing information.

Discrimination of rhythmically-similar languages is acquired at a later age, around 4- to 5-months, and requires a familiarity with at least one of the languages (Bosch & Sebastian-Gallas, 1997; Nazzi, Jusczyk & Johnson, 2000). Thus, American English-learning infants are able to discriminate between English and Dutch, and American and British English, but not German and Dutch. What drives discrimination for older infants remains an open question.

Nazzi et al. (2000) suggest that 4- to 5-month-olds use rhythmic information to discriminate rhythmically-similar languages. Specifically, they suggest infants at this age have learned the specific rhythmic properties of their native language, including the properties that distinguish it from other languages of the same rhythm class. However, our results with adult listeners show that discrimination of rhythmically-similar languages and dialects may depend on listeners' familiarity with the intonational properties of one of the languages. If infants behave as the adults in our experiments do, Nazzi et al.'s results may be due to infants using differences in pitch to distinguish their native language from a foreign language.

Lastly, our study provides a possible method for determining intonational similarity between languages. In our experiment, the presence of pitch information prevented discrimination of American English and German by native American English listeners, even though these languages were distinguishable using rhythmic timing information, while the same was not true for American and Australian English. These results suggest that, contrary to our acoustic measures, listeners may have found the intonation of American English and German more similar than American and Australian English. By replicating and extending this paradigm to other language pairs, it might be possible to generate an intonational similarity matrix for American English or other languages.

2.5 Conclusion

In this paper, we have shown that adults can discriminate between prosodically-similar languages using either rhythmic timing or pitch. Their performance is low, but consistently above chance. However, not all language pairs can be discriminated using a single cue type. Some pairs are discriminable using rhythmic timing, but not pitch, e.g., American English and German, while others are discriminable using pitch, but not rhythmic timing, e.g., American and Australian English. Nevertheless, for rhythmically-similar languages and dialects, American English adult listeners weigh pitch information over rhythmic timing cues, even when pitch is uninformative, and even when it hinders discrimination.

3. Language discrimination by infants: American English and German

3.1 Introduction

The present paper explores infants' ability to discriminate between prosodically similar languages in order to determine what information in the speech signal infants use to perform this task. Previous studies indicate a strong role for prosodic cues, and in particular, researchers have suggested infants use rhythmic cues - the duration and timing of speech segments - to discriminate between languages. In this paper, we seek to show that intonation plays an important role in this task.

A number of researchers have explored infants' ability to discriminate their native language from a non-native language (Bahrick and Pickens, 1988; Mehler et al., 1988; Moon et al., 1993; Bosch and Sebastián-Gallés, 1997; Dehaene-Lambertz and Houston, 1997; Christophe and Morton, 1998; Nazzi et al., 2000; Byers-Heinlein et al., 2010). Researchers have also shown that infants can discriminate between some non-native language pairs (Mehler et al., 1988; Mehler and Christophe, 1995; Christophe and Morton, 1998; Nazzi et al., 1998a; Nazzi et al., 2000; Ramus, 2002b), indicating that infants are capable of discriminating some language pairs, despite having no familiarity with either language. Indeed, the ability to discriminate between languages appears to be at least partially innate and species general, and therefore does not require experience with a language. For example, in addition to newborns (Ramus, 2002b), cotton-top Tamarin monkeys (Ramus et al., 2000) and rats (Toro et al., 2003) have both been shown

to discriminate between Dutch and Japanese. Table 3.1 summarizes the language pairs that have been tested on monolingual infants in language discrimination studies.

It has been claimed that infants are born with the ability to discriminate between prosodically distinct languages, but that infants must undergo some development before they can discriminate between prosodically similar languages (Nazzi et al., 1998; 2000). Thus, as infants age, the ability to distinguish the native language from a non-native language improves. Specifically, although 2-month-old English-learning infants cannot discriminate between British English and Dutch (Christophe and Morton, 1998), 5-month-old English-learning infants can (Nazzi et al., 2000). However, the infants used in these two studies come from different dialect backgrounds. Christophe and Morton (1998) tested British English-learning infants, while Nazzi et al. (2000) tested American English-learning infants.

Differences in dialect have been shown to impact infants' word segmentation (Nazzi et al., 2008) and word recognition abilities (Best et al., 2009; Schmale and Seidl, 2009; Schmale et al., 2010). Also, infants have been shown to be capable of discriminating between some dialects or accents of the same language, but not others (Nazzi et al., 2000; Diehl et al., 2006; Kitamura et al., 2006a; Kitamura et al., 2006b; Kinzler et al., 2007; Butler et al., 2011). Therefore, the difference in dialectal background between the infants tested in Christophe and Morton (1998) and Nazzi et al. (2000) prevents these cases from being treated as anything more than indirect evidence for the development in language discrimination ability.

Infants' Native Language	Age	Tested Languages			Non-native Pair?	Evidence of Discrim.		Source
Canadian English	newborn	Canadian English	&	Tagalog		Yes		Byers-Heinlein et al., 2010
French	newborn	Italian	&	American English	Yes	Yes		Mehler & Christophe 1995
French	newborn	French	&	Russian		Yes		Mehler et al. 1988
Mixed	newborn	French	&	Russian	Yes		No	Mehler et al. 1988
American English	newborn	American English	&	Spanish		Yes		Moon, Cooper & Fifer 1993
Spanish	newborn	Spanish	&	American English		Yes		Moon, Cooper & Fifer 1993
French	newborn	Dutch	&	British English	Yes		No	Nazzi, Bertoncini & Mehler 1998
French	newborn	Japanese	&	British English	Yes	Yes		Nazzi, Bertoncini & Mehler 1998
French	newborn	Dutch	&	Japanese	Yes	Yes		Ramus, 2002
British English	2-months	British English	&	Japanese		Yes		Christophe & Morton 1998
British English	2-months	Dutch	&	Japanese	Yes		No	Christophe & Morton 1998
British English	2-months	British English	&	Dutch			No	Christophe & Morton 1998
British English	2-months	French	&	Japanese	Yes		No	Christophe & Morton 1998
American English	2-months	American English	&	French		Yes		Dehaene-Lambertz & Houston 1998
French	2-months	American English	&	French			No	Dehaene-Lambertz & Houston 1998
American English	2-months	American English	&	Italian		Yes		Mehler et al. 1988
American English	2-months	French	&	Russian	Yes		No	Mehler et al. 1988
Catalan	4-months	Spanish	&	Catalan		Yes		Bosch & Sebastian-Galles, 1997
Spanish	4-months	Spanish	&	Catalan		Yes		Bosch & Sebastian-Galles, 1997
Catalan	4-months	Catalan	&	American English		Yes		Bosch & Sebastian-Galles, 1997
Spanish	4-months	Spanish	&	American English		Yes		Bosch & Sebastian-Galles, 1997
American English	5-months	American English	&	Spanish		Yes		Bahrick and Pickens, 1988
American English	5-months	British English	&	Japanese		Yes		Nazzi, Jusczyk & Johnson, 2000
American English	5-months	Italian	&	Japanese	Yes	Yes		Nazzi, Jusczyk & Johnson, 2000
American English	5-months	Spanish	&	Italian	Yes		No	Nazzi, Jusczyk & Johnson, 2000
American English	5-months	British English	&	Dutch		Yes		Nazzi, Jusczyk & Johnson, 2000
American English	5-months	German	&	Dutch	Yes		No	Nazzi, Jusczyk & Johnson, 2000

Table 3.1. A summary of the language pairs that have been tested on monolingual infants in discrimination/preference studies.

In the first experiment, we sought to directly address whether infants' ability to distinguish their native language from a non-native language improves with age. For this purpose we test 5- and 7-month-old American English-learning infants' ability to discriminate between American English and German.

3.2 Experiment 3.1 – Discrimination with full-cue stimuli

American English-learning 5- and 7-month-olds were tested on their ability to discriminate their native language from a prosodically similar non-native language, German. In this and all following experiments, we use the Headturn Preference Procedure modified for a familiarization-preference task, as in Nazzi et al. (2000; see also Bosch, 1998). In this setup, infants are familiarized to stimuli from one language, and then tested on new stimuli from both the familiar and a novel language. Based on previous research of this type, if infants can discriminate between the languages, they should show a novelty preference. That is, in the test phase, they should orient less towards the language used in the familiarization phase and orient longer towards the novel language.

3.2.1 Methods

Participants. A total of sixteen 5-month-olds (mean age: 153 days; range: 140-167 days; 8 males) and sixteen 7-month-olds (mean age: 217 days; range: 199-251 days; 9 males)

from monolingual American English-speaking homes participated. An additional four 5-month-olds and sixteen 7-month-olds were tested, but excluded due to fussiness (10), stopping to feed (2), caretaker interference or experimental error (2), or because the difference between their average novel and familiar looking times fell more than two standard deviations outside of the mean (6)

Stimuli. The stimuli were modeled on those used in Nazzi et al. (2000) and consisted of 8 American English, and 8 German passages. Each passage was made up of 5 sentences recorded by the same speaker (the sentences were those used in Nazzi et al., 1998). Four female native speakers of each language were recorded in a sound-attenuated booth. Each speaker recorded 10 sentences (2 passages). Efforts were made to minimize voice quality differences within and between languages. Utterances were all recorded as adult-directed speech. Passages were normalized for intensity in Praat to 80 dB. The acoustic properties of the stimuli are given in Table 3.2, including duration and mean f0, and number of syllables.

	English	German
Average Number of Syllables per Passage	89 (6)	91 (6)
Average Duration of Passage	17.2 s (1.0)	17.1 s (2.9)
Mean f0 of Passage	215 Hz (21)	187 Hz (22)
Mean f0 Range	466 Hz (75)	422 Hz (116)

Table 3.2. The acoustic properties of the experimental stimuli, including mean f0, f0 range, duration and number of syllables.

Procedure. The procedure and design were identical to that used by Nazzi et al. (2000). The experiment was conducted using the Headturn Preference Paradigm (HPP;

Kemler Nelson et al., 1995). Each infant was tested individually while seated on her parent's lap in a three-sided pegboard booth. A red light was mounted on either side wall, as well as a loudspeaker, hidden behind the screen so that the infant could not see it. On the center wall, there was one blue light and a camera used to record each session. The camera was connected to a monitor, and the lights and speakers were controlled by a computer, both located outside the booth.

A trial was initiated when the blue light on the center wall began to blink. Once the infant oriented towards the center light, the center light was extinguished and one of the red side lights, chosen at random by the computer, began to blink. When the infant turned towards the red light, the auditory stimulus for that trial began to play and continued until the recording ended, or until the infant failed to maintain orientation towards the light for 3 consecutive seconds. A researcher seated at the computer terminal recorded the duration of the infant's head turns. If the infant looked away for less than 3 s, but then turned back again, the look away time was not included in the orientation time. When the trial ended, or if the infant looked away for more than 3 seconds, the side light was extinguished and the center light began to blink until the infant reoriented towards the center. At that point, one of the side lights was randomly chosen to start blinking, initiating another trial. In order to prevent any influence over the infant's looking time, both the researcher and the infant's caregiver wore sound-attenuating Peltor headphones and listened to music so that they were unaware of the stimuli played during trials.

Design. Each experiment was split into two phases: a familiarization phase and a test phase. The familiarization phase consisted of four passages spoken by two speakers from one of the two languages – for example, four passages of English. Half the infants were familiarized with English sentences and the other half with German sentences. In order to move onto the test phase, infants had to listen to each passage for a total of at least 20 seconds, for a total minimum familiarization time of 80 s.

The test phase consisted of 8 test trials – 4 unheard passages of each language spoken by 2 new speakers per language. The order of presentation of the 8 test trials was randomized for each infant. The average looking times to the familiar and novel languages in the test phase was calculated for each infant and compared statistically.

3.2.2 Results and discussion

Mean orientation times to the familiar and novel language trials in the test phase were calculated for each infant. Overall, of the 16 infants in each age group, nine 5-month-olds and fifteen 7-month-olds had a longer orientation time to the novel language. Across all 5-month-olds, the average orientation times were 10.3 s (SD = 4.5 s) for the novel language and 9.9 s (SD = 3.8 s) for the familiarized language. Across all 7-month-olds, the average orientation times were 9.0 s (SD = 3.1 s) for the novel language and 7.4 s (SD = 3.4 s) for the familiarized language. These results are presented in Figure 3.1.

Additionally, a discrimination ratio was calculated for each infant by dividing the orientation time to novel trials by the total test trial orientation time. Based on Arterberry

and Bornstein (2002), a criterion discrimination ratio significantly above chance was determined. A criterion discrimination ratio (DR) above chance (0.50) was calculated from the formula: Effect size = (DR above chance – DR chance) / Standard deviation, using a moderate effect size of 0.5 (Cohen, 1988) and the standard deviation for all subjects. Six of the sixteen 5-month-olds and eight of the 7-month-olds had a significant novel preference. These numbers were compared using a chi-square test, and there was no significant difference between ages ($X^2(1) = 0.508$; $p = 0.476$).

A three-factor repeated-measures ANOVA with Age (2 levels) and Familiarization Condition (2 levels, English vs. German) as a between-subjects variable and Test Language as the within-subjects variable (2 levels, novel vs. familiar) was used to analyze the results. There was a trending effect of Test Language ($F(1,28) = 3.034$; $p = 0.093$) and Age ($F(1,28) = 3.114$; $p = 0.089$); no other main effects or interactions were significant ($p > 0.1$). The two age groups were also analyzed separately, using two-factor RM-ANOVAs (Familiarization Condition & Test Language as independent variables). For the 5-month-olds, there was no significant main effect of Familiarization Condition ($F(1,14) = 0.163$; $p = 0.693$) or Test Language ($F(1,14) = 0.170$; $p = 0.686$), and no significant interaction between Familiarization and Test Language ($F(1,14) = 0.011$; $p = 0.918$). Thus, 5-month-olds' orientation times to the novel and familiar language did not differ significantly. For the 7-month-olds, there was a significant main effect of Test Language ($F(1, 14) = 22.742$; $p < 0.001$). There was no effect of Familiarization Condition ($F(1,14) = 0.914$; $p = 0.355$), and no interaction ($F(1,14) = 0.078$; $p = 0.784$).

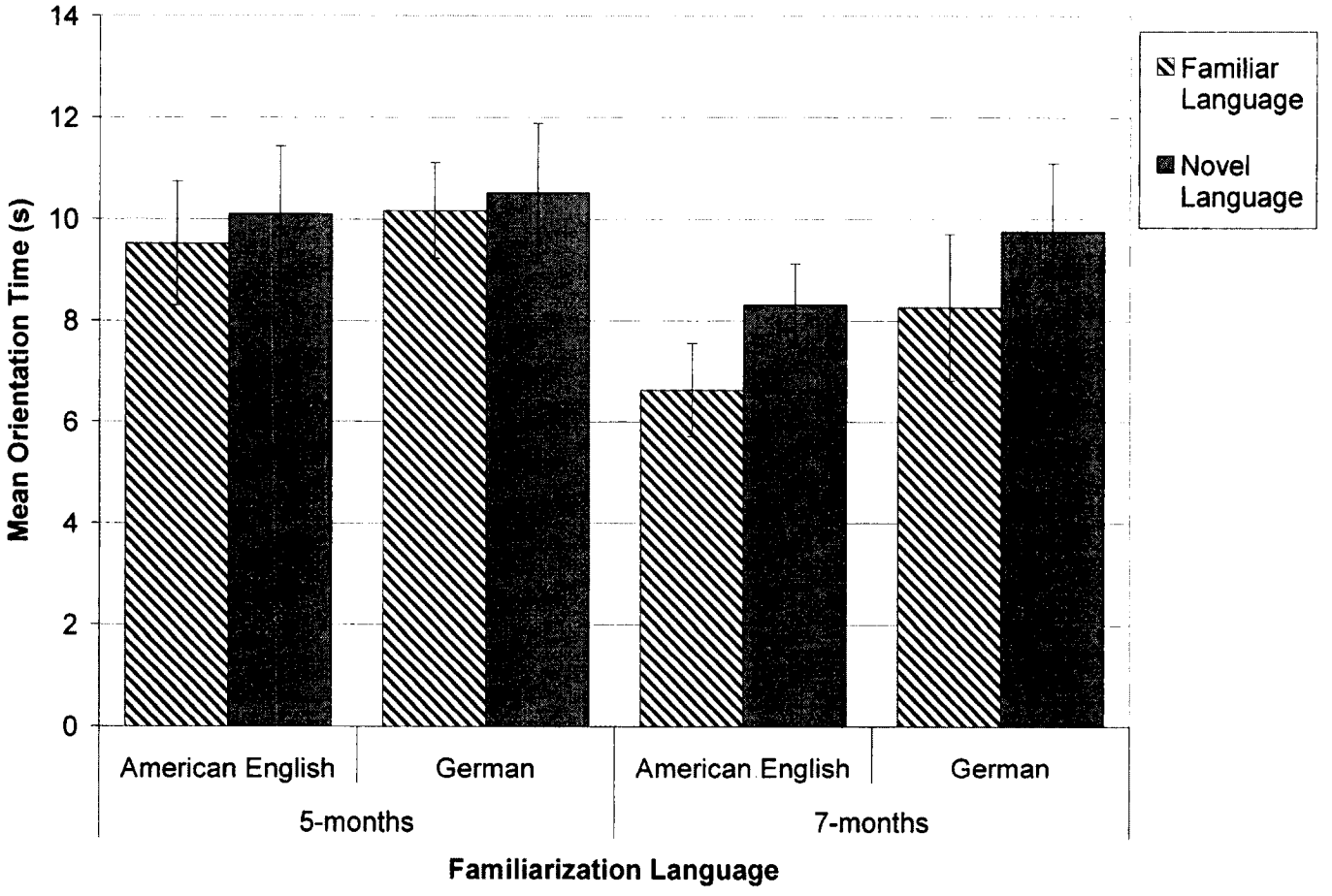

Figure 3.1. Mean orientation times (and standard error bars) of 5- and 7-month-olds to languages used in the test phase in Experiment 3.1, using full-cue stimuli. Orientation times for each age are separated by familiarization language.

Thus, 7-month-olds showed significantly longer orientation times to the novel language over the familiarized language.

Results from Experiment 3.1 show that American English-learning 7-, but not 5-month-olds, were able to successfully discriminate their native language from a prosodically similar non-native language, German, indicating this ability develops between 5- and 7-months. The current study is the first to directly show that infants' language discrimination ability develops with age using identical testing materials and infants from the same linguistic background.

However, these results are surprising considering previous research showing that American English-learning 5-month-olds can discriminate between British English and Dutch, and between British English and their native dialect (Nazzi et al., 2000). Because 5-month-olds can discriminate between these language/dialect pairs, it was expected that they would be able to discriminate between American English and German, as well. That they can't discriminate these languages until 7-months, suggests that the differences between American English and German are smaller than those between British English and Dutch, at least on the acoustic dimensions attended to by infants. It seems that some language pairs are perceptually more similar than others, requiring more experience before infants are able to discriminate them.

3.3 Experiment 3.2 – Discrimination with low-pass filtered stimuli

In the previous experiment, we showed that American English-learning infants' ability to discriminate between their native language and German develops between 5- and 7-months of age. What information in the speech signal are 7-month-olds attending to, and what have they learned about language that allows them to discriminate languages that they were previously unable to distinguish?

Several studies have shown infants can discriminate languages using low-pass filtered speech, suggesting that infants rely on prosodic information. These include studies on newborns (Mehler et al., 1988; Nazzi et al., 1998a; Byers-Heinlein et al., 2010), 2-month-olds (Mehler et al., 1988; Dehaene-Lambertz & Houston, 1998), and 4-month-olds (Bosch and Sebastián-Gallés, 1997). Low-pass filtering eliminates most of the segmental information of speech stimuli while preserving prosody. Additional researchers have shown that infants can rely on prosodic cues by using other methods. Ramus (2002) showed newborns could discriminate language stimuli that had been re-synthesized, a method which removes all segmental information, but preserves prosodic information and some phonotactic information. In a study with 2-month-olds, Dehaene-Lambertz & Houston (1998) tested for discrimination using (a) unaltered, (b) filtered and (c) scrambled sentences. Scrambling disrupts the prosodic coherence of the overall sentence, but preserves segmental information and word-level prosody. While infants could successfully discriminate the unaltered and filtered stimuli, they were unable to discriminate the scrambled stimuli. Therefore, it seems likely that infants between birth

and 4-months are relying on utterance-level prosodic information to discriminate languages.

Nazzi et al. (2000) have also argued that it is unlikely that infants are attending to broad types of segmental information, like phonotactic information, or picking out word forms. They showed that American English-learning 5-month-olds could discriminate between American and British English. In their study, the same passages were used for the American and British stimuli. That is, the same words and phonotactics were used in the two sets of stimuli, yet infants were still able to discriminate between them. However, infants were only tested on unmodified natural stimuli in that study. Thus, it is possible they were relying on more narrow sources of segmental information to discriminate American and British English, like dialectal differences in phonetic realizations of the same phonemes.

It is well known that infants, even newborns, are capable of perceiving a number of phonetic contrasts, including both consonantal and vocalic contrasts, regardless of whether or not the contrasts are in their native language, and that this ability reduces with age (Eimas et al., 1971; Morse, 1972; Trehub, 1973; 1976; Werker et al., 1981; Werker and Tees, 1984). Evidence suggests this change is due to infants' growing distributional awareness of the segments and contrasts important to their native language (Maye et al., 2002; Maye et al., 2008; Yoshida et al., 2010). With these abilities, infants could perceive the different segments used in two different language samples, track their distributional frequency, and use that information to discriminate between them. Such a strategy for language discrimination has been suggested by Bosch and Sebastián-Gallés

(2001) to account for monolingual and bilingual infants' ability to distinguish Catalan and Spanish.

Infants do not home into their native language in this way until 8- to 10-months of age, at least for consonantal contrasts (Werker and Tees, 1984; Kuhl et al., 1992; Polka and Werker, 1994). However, there is some evidence that the perception of vowels reorganizes earlier, between 4- and 6-months (Kuhl et al., 1992; Polka and Werker, 1994; Polka and Bohn, 1996), which is the age examined in some language discrimination studies (Bosch and Sebastián-Gallés, 1997; Nazzi et al., 2000; Bosch and Sebastián-Gallés, 2001), including the current study. Therefore, it is possible that 7-month-olds tested in Experiment 3.1 are attending to differences in vowels, or using other types of segmental information when discriminating languages.

In Experiment 3.2 we tested the hypothesis that 7-month-old American English-learning infants use segmental information to distinguish English and German. For this we used low-pass filtering to remove most segmental information from the stimuli. If the 7-month-olds in Experiment 3.1 were relying on segmental information to discriminate between American English and German, then they should fail to discriminate the two languages when they have been low-pass filtered. If, instead, infants are relying on prosodic information, then discrimination should still be observed.

3.3.1 Methods

Participants. Sixteen American 7-month-olds (mean age: 212 days; range: 202-225 days; 4 males) from monolingual English-speaking homes participated. An additional 8 infants were tested but their data were not included because they failed to complete the experiment due to fussiness (n=3), or because the difference between their average novel and familiar looking times fell more than two standard deviations outside of the mean (5).

Stimuli. The English and German stimuli from Experiment 3.1 were low-pass filtered using Praat at a frequency cut-off of 400 Hz, with 50 Hz smoothing.

Procedure and Design. The procedure and design used in this experiment were identical to Experiment 3.1.

3.3.2 Results and discussion

Mean orientation times to the familiar and novel language trials in the test phase were calculated for each infant. Overall, 12 of the 16 infants had a longer orientation time to the novel language. Across all infants, the average orientation times were 7.7 s (SD = 2.6 s) for the novel language and 6.6 s (SD = 2.0 s) for the familiarized language, as shown in Figure 2. Additionally, a discrimination ratio was calculated for each infant and the number of infants showing a significant novel preference was calculated. Seven of the 16 infants showed a significant novel preference. This was compared to the 7-month-old

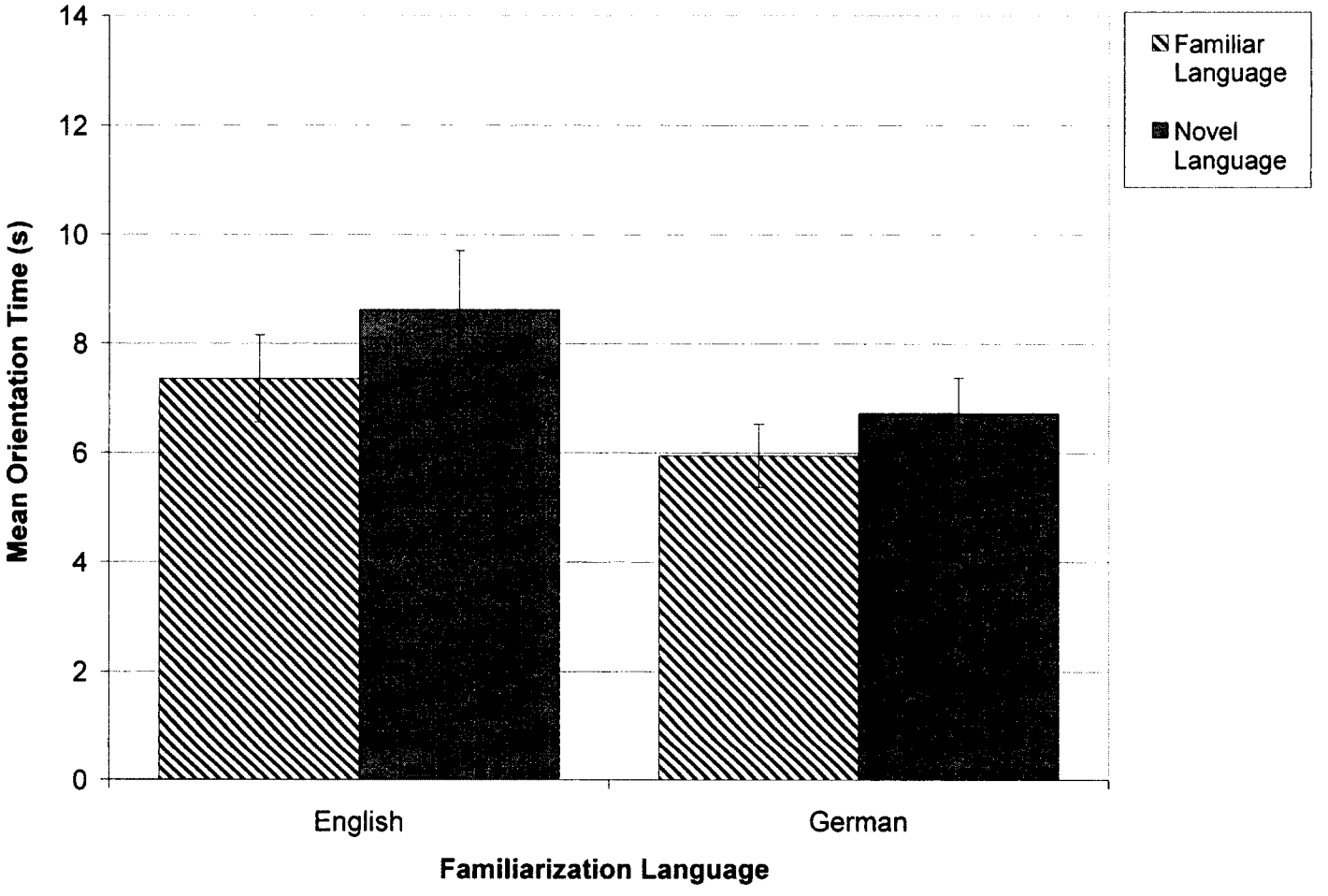

Figure 3.2. Mean orientation times (and standard error bars) of 7-month-olds for languages used in the test phase in Experiment 3.2, using low-pass filtered stimuli. Orientation times are separated by familiarization language.

group from Experiment 3.1 using a chi-square test, but there was no difference between groups ($X^2(1) = 0.125; p = 0.724$).

A two-factor repeated-measures ANOVA with Familiarization Condition (2 levels, English vs. German) as a between-subjects variable and Test Language (2 levels, novel vs. familiar) as the within-subjects variable showed a significant main effect of Test Language ($F(1, 14) = 8.384$; $p = 0.010$). There was no effect of Familiarization Language ($F(1,14) = 2.355$; $p = 0.147$), and no interaction between Familiarization and Test Language ($F(1,14) = 0.534$; $p = 0.477$). Thus, infants showed significantly longer orientation times to the novel language compared to the familiarized language, indicating that American English-learning 7-month-olds were still able to discriminate between English and German, even when the stimuli had been low-pass filtered.

To compare the results of the current experiment to those from the 7-month-olds tested in Experiment 3.1, a three-factor RM-ANOVA (Experiment x Familiarization Condition x Test Language) was used. There was a significant main effect of Test Language ($F(1,28) = 29.759$, $p < 0.001$), as both groups showed a preference for the novel language during the test phase. However, there was no significant effect of Experiment ($F(1,28) = 1.239$; $p = 0.275$) or any interaction between Experiment and Test Language ($F(1,28) = 1.423$; $p = 0.243$). Thus, infants listening to low-pass filtered stimuli did not behave differently from those listening to unmodified, full cue stimuli.

These results confirm that infants can rely on prosodic information alone when discriminating between languages, even for prosodically similar languages like English and German. Segmental information is not necessary for discrimination.

3.4 Experiment 3.3 – Discrimination with re-synthesized intonation stimuli

The prosodic information available in the speech signal after low pass filtering includes both intonation and linguistic rhythm. Languages have frequently been classified in terms of their rhythm, since Pike (1945) and Abercrombie (1967) as either "stress-timed" or "syllable-timed," or if a more continuous classification scheme is assumed, somewhere in between. It has also been suggested that membership in each of these classes affects the way a language is processed by its native speakers (Mehler et al., 1981; Cutler et al., 1986; Cutler and Norris, 1988; Sebastian-Galles et al., 1992; Bradley et al., 1993; Otake et al., 1993; Pallier et al., 1993).

The idea that language rhythm helps shape processing strategies has influenced researchers' view of infant discrimination abilities. Nazzi et al. (1998) observed that young infants (newborns to 2-month-olds) only seem to discriminate between languages when they come from different rhythm classes. This led them to propose the *rhythm hypothesis*, which suggests that infants "extract rhythmic properties from speech and sort sentences into a small number of classes or sets based on rhythmic, timing properties" (Nazzi et al., 1998; see also Mehler et al., 1996, for a similar proposal).

However, older infants have also been observed to discriminate between languages from *within* the same rhythm class. Two-month-old English-learning infants cannot discriminate between British English and Dutch, two stress-timed languages (Christophe and Morton, 1998), but 5-month-old English-learning infants can (Nazzi et al., 2000). Similarly, Spanish- and Catalan-learning 4-month-olds can discriminate

between Spanish and Catalan, two syllable-timed languages (Bosch and Sebastián-Gallés, 1997).

Nazzi et al. (2000) offer a revised version of the *rhythm hypothesis* to explain this development in behavior – the *native-language acquisition hypothesis.* This hypothesis asserts that while younger infants can only discriminate languages based on broad classifications of rhythm, 4- to 5-month-olds have additionally learned to focus their attention on the specific rhythmic properties that distinguish their native language from other languages within the same rhythm class. For example, while English-learning infants at this age discriminate between English & Dutch, they cannot discriminate between Dutch & German, two non-native stress-timed languages, or between Italian-Spanish, two non-native syllable-timed languages (Nazzi et al., 2000).

Research into linguistic rhythm has suggested that rhythm arises from the phonological properties of a language, such as the phonotactic permissiveness of consonant clusters, the presence or absence of contrastive vowel length and vowel reduction (Dauer, 1983). This line of thought has led to the development of a variety of rhythm metrics intended to categorize languages into classes using measurements made on the duration of segmental intervals (Ramus et al., 1999; Grabe and Low, 2002; Wagner and Dellwo, 2004; White and Mattys, 2007). Since then, a number of researchers have equated linguistic rhythm with segmental duration and timing, the target of these rhythm metrics (Ramus and Mehler, 1999; Guasti, 2002; Ramus, 2002b; Byers-Heinlein et al., 2010).

Although these various metrics have been shown to successfully differentiate between prototypical languages from different rhythm classes on controlled speech materials, they are less successful with uncontrolled materials and on non-prototypical languages (Ramus, 2002a; Arvaniti, 2009; Wiget et al., 2010). Other researchers have criticized these metrics as having moved away from the original notions of speech rhythm, and even of Dauer's analysis, on which they were originally based (Kohler, 2009), and have pointed out that they are more measures of speech timing than rhythm (Arvaniti, 2009).

In fact, there may not be such a thing as neatly defined rhythm classes at all. Indeed, the notion that languages could be labeled exclusively as either stress-timed or syllable-timed originated after the introduction of those terms (Kohler, 2009). Languages were originally described as having both qualities, to differing extents (Pike, 1945). As a consequence, numerous languages have proved resistant to being neatly classified into these groups (Dauer, 1983; Ramus, 2002a; Arvaniti, 2009). Nevertheless, segmental duration and timing is an important characteristic of languages, and adult listeners have been shown to be sensitive to this information when discriminating languages (Ramus and Mehler, 1999; Chapter 2).

Besides segmental duration and timing, there is another aspect of prosody infants may be attending to when discriminating languages – intonation, or pitch. All languages make use of pitch. Some languages, like Korean or French, use pitch to mark word edges (Jun and Fougeron, 2000; Jun, 2005; Kim & Cho, 2009). However, in stress languages like English and German, pitch is used to mark specific syllables within a word, and can

be one of several possible contours depending on the context in which the word is produced (Pierrehumbert, 1980; Grice et al., 2005). Regardless of language-specific differences in its instantiation, all known languages use pitch to mark the ends of large phrases (Jun, 2005) and pitch contours over the whole sentence consist of interpolated pitch between sentence-internal and sentence-final marking (Ladd, 1996).

Because pitch is a universally important component of the speech signal, it is reasonable to think that listeners would make use of this information when discriminating between languages. Both infants and adults are certainly sensitive to pitch. Research has also shown that cross-linguistically, adults use pitch information to segment words from running speech (Reitveld, 1980; Bagou et al., 2002; Welby, 2003; Kim, 2004; Tyler et al., 2006; Tyler and Cutler, 2009; Choi and Mazuka, to appear). Indeed, several studies have shown that adults are capable of discriminating between languages using only pitch cues (Ramus and Mehler, 1999; Komatsu et al., 2004), including prosodically similar languages like English-Dutch (Willems, 1982; de Pijper, 1983), and even dialects like American-Australian English (Chapter 2). Based on this research, it is likely that pitch information factors into the discrimination ability of infants. Indeed, newborns are better able to discriminate languages from separate rhythm classes when pitch differences are preserved rather than degraded (Ramus, 2002).

Although they suggest that infants use rhythmic information to discriminate between their native language and prosodically similar non-native languages, there is some evidence based on Nazzi et al.'s (2000) results that infants could have used intonation to distinguish between languages in that study. Specifically, they showed that

American English-learning 5-month-olds could discriminate between American and British English. Although they report that their American and British English stimuli were closely matched in pitch, "the British English utterances had a greater proportion of higher sentence-initial pitches and a greater proportion of terminal rises than the American-English utterances" (Nazzi et al., 2000).

Results from Experiment 3.2 showed that infants could still discriminate English and German sentences when segmental information had been removed from the stimuli (or at least severely degraded), indicating that 7-month-olds can rely on prosodic information alone to discriminate English and German. Experiment 3.3 was designed to test whether 7-month-olds are using pitch information to distinguish English and German. For this purpose, we used modified stimuli with re-synthesized pitch contours. A simple undulating pitch contour was added to the American English and German stimuli so that all sentences in both languages had the same unfamiliar pitch pattern. This was done instead of simply removing the pitch contour, leaving a monotone pitch, because it was thought monotone intonation would be too boring and increase attrition. If infants are relying on intonation information, then they should fail to discriminate between the re-synthesized English and German passages. However, if infants are relying on rhythmic durational and timing information to discriminate English and German, then discrimination should still be observed in Experiment 3.3.

3.4.1 Methods

Participants. Sixteen American 7-month-olds (mean age: 216 days; range: 198-231 days; 5 males) from monolingual English-speaking homes participated. An additional 7 infants were tested but their data were not included because they failed to complete the experiment due to fussiness (3), or because the difference between their average novel and familiar looking times fell more than two standard deviations outside of the mean (4).

Stimuli. The English and German stimuli from Experiment 1 were modified in Praat for use in the current experiment. The original pitch contours of the passages were removed and replaced with an artificial contour consisting of a repeating sine-wave with a period of 400 ms, a minimum pitch of 185 Hz and a maximum pitch of 245 Hz. All passages, in both English and German, received this contour.

Procedure and Design. The procedure and design used in this experiment were identical to Experiment 3.1.

3.4.2 Results and discussion

Mean orientation times to the familiar and novel language trials in the test phase were calculated for each infant. Four of the 16 infants had a longer orientation time to the novel language. Across all infants, the average orientation times were 7.7 s (SD = 2.7 s) for the novel language and 8.6 s (SD = 2.9 s) for the familiarized language, as shown in Figure 3.3. Additionally, a discrimination ratio was calculated for each infant and the

number of infants showing a significant novelty preference was calculated. Only 2 of the 16 infants showed a significant novel preference. This was compared to the 7-month-old group from Experiment 3.1 using a chi-square test, and there was a significant difference between groups ($X^2(1) = 5.236$; $p = 0.022$).

A two-factor repeated-measures ANOVA with Familiarization Condition (2 levels, English vs. German) as a between-subjects variable and Test Language (2 levels, novel vs. familiar) as the within-subjects variable was conducted. There was no significant main effect of Familiarization Language ($F(1,14) = 2.366$; $p = 0.146$) and a trending effect of Test Language ($F(1, 14) = 3.588$; $p = 0.079$). However, there was a significant interaction between Familiarization and Test Language ($F(1,14) = 7.588$; $p = 0.016$). Post-hoc paired t-tests were used to probe this interaction by looking at the test language preference for each familiarization group. Infants familiarized to re-synthesized American English stimuli show no significant preference during the test phase ($t(7) = 0.630$). However, infants familiarized to re-synthesized German stimuli showed higher mean orientation times to the familiar language, German, during the test trials ($t(7) = 3.182$, $p = 0.015$).

A three-factor RM-ANOVA (Experiment x Familiarization Condition x Test Language) was used to compare the results of the current experiment to the 7-month-old group in Experiment 3.1. There was no significant main effect of Experiment ($F(1,28) = 0.003$), but there was an interaction between Experiment and Test Language ($F(1,28) =$

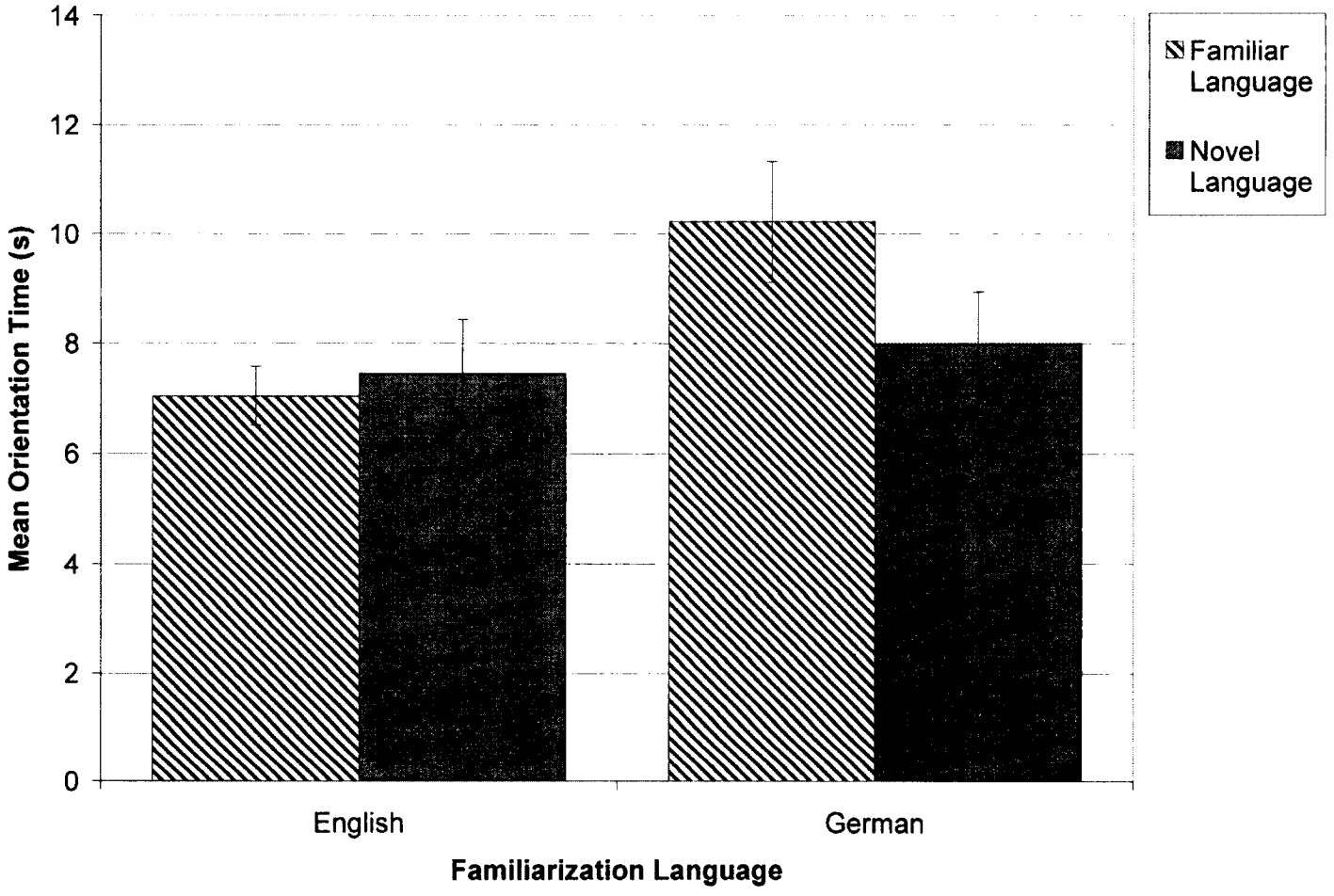

Figure 3.3. Mean orientation times (and standard error bars) of 7-month-olds for languages used in the test phase in Experiment 3.3, using stimuli re-synthesized with artificial intonation. Orientation times are separated by familiarization language.

18.258; $p < 0.001$). Infants in the first experiment showed a preference for the novel language during the test trials, indicating they were able to discriminate between English and German. But this was not the case for the current experiment.

Because of the lack of a clear novelty preference across infants in Experiment 3.3, we can conclude that American English-learning 7-month-olds were unable to discriminate between American English and German once the stimuli had been re-synthesized to eliminate intonational information. Thus, intonation was necessary for infants to discriminate English and German in the Experiments 3.1 and 3.2.

It is possible that infants failed to discriminate between the re-synthesized stimuli, not because the missing intonational cues eliminated important distinctions between the languages, but simply because the re-synthesis made the stimuli sound unnatural or disturbing. This seems unlikely for several reasons. First, the stimuli did not sound overly unnatural or bizarre to adult ears. Secondly, fuss-out rates were no higher in this experiment than in the previous experiments (30% vs. 33% in Experiment 3.2 and 50% in Experiment 3.1), and finally, average orientation times to the re-synthesized speech were not significantly different than orientation times to the stimuli used in Experiments 3.1 or 3.2 (8.2 s, SD = 2.5 vs. 8.2 s, SD = 3.2 for Experiment 3.1; $t(30) = 0.055$; $p = 0.956$; and 7.2 s, SD = 2.4 for Experiment 3.2; $t(30) = 1.208$; $p = 0.237$). Therefore, it is more likely that infants failed to discriminate in the current experiment because they lacked access to intonational cues.

3.5 General discussion

Following the footsteps of Nazzi et al. (2000), in Experiment 3.1 we tested American English learning 5- and 7-month-olds on their ability to discriminate between two prosodically similar languages – their native language, American English, and a non-native language, German. The younger infants were unable to discriminate the two languages, but the older infants were successful, indicating that this ability develops between 5- and 7-months of age. These results are in line with previous research indicating that language discrimination ability develops over the first year of life (Christophe & Morton, 1998; Nazzi et al., 2000), but provide the first direct evidence for this claim.

These results contrast with the strong interpretation from previous research that 5-month-olds can discriminate their native language from other languages within the native rhythm class. To be fair, no researcher has explicitly claimed that infants should be capable of discriminating their native language from all other rhythmically similar languages by 5-months of age. Infants may find American English and German more similar than other language pairs. It would therefore take them more time to learn to distinguish this pair than other language pairs.

Next, in Experiment 3.2, we tested whether 7-month-olds rely on segmental information to discriminate between English and German. Numerous researchers have shown that segmental information is not necessary for infants to discriminate between languages (Mehler et al., 1988; Bosch & Sebastian-Galles, 1997; Nazzi, Bertoncini &

Mehler, 1998; Dehaene-Lambertz & Houston, 1998; Ramus, 2002), though this research was conducted on younger infants, between 0- and 4-months, and on prosodically dissimilar languages. It is possible that by 7-months, infants are attending to and using segmental information to discriminate between languages. However, with infants' success in discriminating low-pass filtered American English and German in Experiment 2, we have shown that segmental information is also not necessary for 7-month-olds to discriminate between their native and a non-native language, even for prosodically similar languages like English and German. Additionally, recall that in Experiment 3.3, infants had access to complete segmental information, only the intonation was modified, yet 7-month-olds failed to discriminate English and German in this experiment. Thus, we show that access to segmental information is also not sufficient to cue language discrimination. Together, these results support the hypothesis put forth in Nazzi et al. (2000) that the increase in infants' language discrimination ability is due to an increase in ability to perform fine-grained analyses of the prosodic properties of their native language.

There are two main prosodic cues infants could be using to discriminate between languages – rhythm or intonation. In Experiment 3.3, when exposed to stimuli that had been re-synthesized with an artificial pitch contour, eliminating any natural pitch information or differences between the languages, but leaving rhythmic and segmental cues intact, 7-month-olds failed to discriminate American English and German. Thus, intonation is necessary for infants to discriminate between American English and German; rhythmic information is not sufficient to support their discrimination.

It is not surprising that intonation plays an important role in infants' language discrimination. Infants' ability to hear and perceive pitch becomes adult-like relatively early. At 3-months, frequency resolution around 1,000 Hz is adult-like, and resolution at higher frequencies matures by 6-months (Schneider et al., 1990; Spetner and Olsho, 1990). Seven- to 8-month-olds perceive frequency harmonics as equivalent to the fundamental frequency, as adults do (Clarkson and Clifton, 1985; Montgomery and Clarkson, 1997). In the linguistic domain, French newborns are able to discriminate between lists of Japanese words containing segmentally identical tokens which differ only in their pitch contour (Nazzi et al., 1998b). American English-learning 2- to 3-month-olds have been shown to discriminate between rising and falling pitch contours on syllables (Karzon and Nicholas, 1989), and American English-learning infants between 5- to 11-months of age can discriminate between words which vary only in the pitch on the final syllable, even if the difference in f0 is as low as 5 or 10 Hz (Bull et al., 1985).

There is also evidence that infants use pitch in processing speech. Infants prefer infant-directed speech over adult-directed speech because of the differences in pitch patterns, not because of differences in amplitude or duration (Fernald and Kuhl, 1987). Further, 6-month-olds prefer to listen to sentences with pauses coincident with clause boundaries, indicating a preference for continuous over discontinuous pitch contours (Hirsh-Pasek et al., 1987). In fact, 6-month-old American English-learning infants weight pitch over other cues that mark clause boundaries, such as pauses or pre-boundary-lengthening (although they require at least one of these secondary cues along with pitch to successfully segment clausal units; Seidl, 2007). This is a pattern that

develops between 4- and 6-months, as 4-month-olds show no preference for pitch cues in this task (Seidl and Cristià, 2008). This is the same age infants are thought to begin developing the ability to discriminate between certain language pairs that were not distinguishable at birth (Nazzi et al., 2000).

In contrast, it would be difficult for infants to rely on only rhythmic differences to discriminate between rhythmically similar languages, like American English and German. Research on classifying languages using differences in segmental duration and timing have shown that inter-speaker variability within a single language can often be as big as cross-language variability, preventing reliable classification, even for languages that are traditionally considered to be from different rhythm classes (Arvaniti, 2009; Wiget et al., 2010; Loukina et al., 2011). Indeed, results from the current experiment indicate that infants cannot use this type of information to discriminate between English and German.

This result suggests that the infants tested in Nazzi et al. (2000) were likely discriminating rhythmically similar languages based on differences in pitch patterns, rather than rhythmic differences. It may also explain why infants were more successful in discriminating rhythmically-similar languages only when one of the languages was familiar to them. Specifically, infants in that study could discriminate between British English and Dutch, as well as between American and British English, but failed to discriminate between German and Dutch. Previous research on adults' discrimination ability has shown that adults can use pitch information to discriminate between languages, including intonationally similar languages (Willems, 1982; de Pijper, 1983;

(Ramus and Mehler, 1999; Komatsu et al., 2004). However, there is also evidence indicating that the ability to use pitch cues to discriminate requires prior familiarity with at least one of the languages (Barkat et al., 1999; Ramus and Mehler, 1999). Thus, infants in Nazzi et al. (2000) may have failed to discriminate between German and Dutch, not because they were unable to distinguish between the different rhythmic patterns, but because the intonational systems of both languages were unfamiliar to them.

The fact that infants attend to intonation when discriminating languages has implications beyond Nazzi et al.'s (2000) study. According to the rhythm hypothesis (Nazzi et al., 1998), upon birth, infants should be able to discriminate between any two languages that belong to different rhythm classes. However, as mentioned in the first section, many languages have proven difficult to classify in this way (Dauer, 1983; Ramus, 2002b; Arvaniti, 2009). Indeed, there have been several counter-examples to the rhythm hypothesis in the developmental literature. For instance, English-learning 2-month-olds fail to discriminate between French and Russian, a supposed syllable-timed and stress-timed language, respectively (Mehler et al., 1998) or between French and Japanese, the latter of which has been classified as mora-timed (Christophe & Morton, 1998; although this may be due to methodological insensitivities, see Nazzi et al., 2000). Rather than discriminating languages based only on differences in rhythm, it seems more likely that young infants are listening to intonation, or possibly both rhythmic and intonational information. If so, languages that are difficult to classify rhythmically might still be distinguishable if they have large differences in intonation. Similarly, cases where infants fail to discriminate between rhythmically distinct languages could be

because the intonation systems of the tested languages are too similar or unfamiliar. Testing this provides an avenue for future research.

Currently, it is difficult to predict which languages infants would find intonationally distinct. Although there have been attempts to acoustically define and describe rhythm and rhythmic differences between languages (i.e., the various rhythm metrics; Ramus et al., 1999; Grabe & Low, 2002; Wagner & Dellwo, 2004; White & Mattys, 2007), there have been no similar major efforts to compare intonation cross-linguistically, at least at a phonetic level. The intonation systems of many languages have been described from a phonological perspective, though, often using the ToBI framework and autosegmental-metrical phonology (e.g., Pierrehumbert, 1980, Ladd, 1996). This research has been used to develop a basic intonational typology – whether languages use intonation to mark word-edges or word-heads, for example (Jun, 2005). However, infants would almost certainly attend to information beyond what is captured by these analyses, such as how often specific pitch contours appear at different salient positions within the utterance. Clearly, additional research into cross-linguistic intonational patterns is required before infants' discrimination ability with a specific language pair can be easily predicted.

In conclusion, the current investigation confirms that the ability to discriminate between languages develops with age, allowing for the emergence of the ability to discriminate the native language from a prosodically similar non-native language. This development in discrimination ability is not due to an attention to segmental information. Rather, infants can rely solely on prosodic information to discriminate. In particular,

pitch information is necessary for infants to discriminate between prosodically similar languages.

4. Language discrimination by infants: Additional language pairs

4.1 Introduction

In the second chapter of this thesis, American English-speaking adults were shown to be able to discriminate their native language from a prosodically similar non-native language, German, as well as from a non-native dialect, Australian English, using only prosodic cues. It was shown that adults could use both rhythm and intonation to discriminate languages/dialects, but intonational cues were always weighted above rhythmic cues. To discriminate between American and Australian English, American English-speaking adults *required* the presence of intonational information; rhythmic differences between the dialects alone were not sufficient to support discrimination. In contrast, adults' reliance on intonation *hindered* their ability to discriminate between American English and German using rhythm alone. When the distracting intonational information was removed from the stimuli, adults could rely on rhythmic differences to discriminate between the languages.

In the third chapter of the thesis, American English-learning infants were also shown to be capable of discriminating between American English and German using only prosodic cues. For infants, intonation cues were necessary to discriminate American English and German; rhythmic differences alone between the languages were not sufficient to support discrimination. In this chapter, infants' ability to discriminate languages is tested on a number of additional language pairs – with little success. The

implications that this series of null and unexpected results have on infants' language discrimination abilities and infant testing methodologies are discussed in the following sections.

4.2 Experiment 4.1 - American and Australian English

From a practical standpoint, the most obvious language pair to test on infants following the previous experiments described in this thesis is American and Australian English, as adults were tested on both that language pair and American English-German in Chapter 2, and infants were tested only on American English-German in Chapter 3.

Previous research indicates that infants are capable of discriminating between different dialects and accents of their native language. Nazzi et al. (2000), using the Headturn Preference Paradigm (HPP) in a familiarization-preference design, showed that American English-learning 5-month-olds could discriminate between their native dialect and British English. As mentioned in Chapter 1, they claimed that infants at this age were able to construct a detailed representation of the rhythmic properties of their native language, to which they could compare heard utterances and identify them as native or non-native. This representation was detailed enough that infants could discriminate their native language from rhythmically similar non-native languages or dialects, like American and British English, but also British English and Dutch. Rhythmically similar utterances judged as outside the native language were seemingly grouped together into a broad 'non-native' category. For example, 5-month-olds could not discriminate between

German and Dutch (Nazzi et al., 2000). Similarly, using the same sentence script and testing methodology, Butler et al. (2011) showed that South-West English-learning 5- and 7-month-olds could discriminate between their native dialect and a non-native dialect, Welsh English, but were unable to discriminate between two non-native dialects, Welsh and Scottish English.

In this experiment, American English-learning infants are tested on their ability to discriminate between their native dialect and a non-native dialect, Australian English. This language pair has been tested before in a series of language preference experiments by Diehl, Kitamura, Panneton and colleagues (Diehl et al., 2006; Kitamura et al., 2006a; Kitamura et al., 2006b). Four female speakers of each dialect recorded a set of five sentences in infant-directed speech. Three passages, or trial stimulus items, were then constructed for each dialect consisting of one token of each sentence, but each sentence was spoken by a different speaker. Thus, on each trial, infants heard multiple sentences by multiple speakers. Only the dialect was constant. They found that 6-month-old American English-learning infants and 3-month-old Australian English-learning infants preferred to listen to Australian English over American English, indicating that they could discriminate between the two dialects.

However, they also found that dialect preference developed with age. In their studies, they showed that while American 6-month-olds and Australian 3-month-olds showed a preference for Australian English, older infants – American 8-month-olds and Australian 6-month-olds – showed no such preference. When their discrimination ability was tested directly, Australian 6-month-olds failed to discriminate between the two

dialects (Kitamura et al., 2006), indicating a loss in discrimination ability between 3- and 6-months. For this dialect pair, American English-learning infants lost the ability to discriminate at a later age, between 6- and 8-months. Thus, unlike infants' language discrimination abilities that improve with age, particularly those that allow infants to discriminate prosodically similar languages, infants' ability to discriminate dialects seem to decline with increasing age.

In the current experiment, American English-learning infants from three age groups – 5-, 7- and 9-months – were tested on their ability to discriminate between American and Australian English. Based on Nazzi et al. (2000) and Butler et al.'s (2011) results, 5-month-olds were predicted to show discrimination. Based on Diehl et al. (2006) and Kitamura et al. (2006, 2006a) 9-month-olds were expected to fail. Seven-month-olds should serve as an intermediate case, where, according to previous research, the ability to discriminate these dialects is in a state of decline.

4.2.1 Methods

Participants. Infants from three age groups were tested. Twenty-two American 5-month-olds (16 males), fourteen American 7-month-olds (7 males), and eight American 9-month-olds (3 males) from monolingual English-speaking homes participated. The mean age was 154 days (range = 138 to 168 days) for the 5-month group, 213 days (range = 200 to 225 days) for the 7-month group and 269 days (range = 259 to 279 days) for the 9-month group. An additional seven 5-month-olds, five 7-month-olds and eight 9-month-

olds were tested but their data were not included because they failed to complete the experiment due to fussiness and crying (17) or lack of interest (3).

Stimuli. The stimuli were modeled on those used in Nazzi et al. (2000) and consisted of 8 American English and 8 Australian English passages, constructed from the sentences used in the experiments in Chapter 2. To repeat, those sentences consisted of 39 sentences taken from Nazzi et al. (1998) that were recorded by eight female speakers of American English and eight female speakers of Australian English in a sound-attenuated booth or quiet room at a sampling rate of 22050 Hz. Efforts were made to minimize voice quality differences within and between languages. Utterances were all recorded as adult-directed speech.

	American English	Australian English
Average Number of Syllables per Passage	89 (6)	90 (5)
Average Duration	17.2 s (1.0)	18.6 s (0.6)
Mean f0	215 Hz (21)	236 Hz (23)
Mean f0 Range	466 Hz (75)	396 Hz (119)

Table 4.1. Acoustic properties of the American and Australian English passages.

Each passage was made up of 5 sentences recorded by the same speaker. Only 4 of the original 8 speakers were used. Thus, each speaker recorded 10 sentences (2 passages). Passages were normalized for intensity in Praat to 80 dB. The acoustic properties of the stimuli are given in Table 4.1, including number of syllables, duration, mean f0, and f0 range.

Procedure. The procedure and design were identical to that used by Nazzi et al. (2000), and in Chapter 3. The experiment was conducted using the Headturn Preference Paradigm (HPP; Kemler Nelson et al., 1995). Each infant was tested individually while seated on her parent's lap in a three-sided pegboard booth. A red light was mounted on either side wall, as well as a loudspeaker, hidden behind the screen so that the infant could not see it. On the center wall, there was one blue light and a camera used to record each session. The camera was connected to a monitor, and the lights and speakers were controlled by a computer, both located outside the booth.

A trial was initiated when the blue light on the center wall began to blink. Once the infant oriented towards the center light, the center light was extinguished and one of the red side lights, chosen at random by the computer, began to blink. When the infant turned towards the red light, the auditory stimulus for that trial began to play and continued until the recording ended, or until the infant failed to maintain orientation towards the light for 3 consecutive seconds. A researcher seated at the computer terminal recorded the duration of the infant's head turns. If the infant looked away for less than 3 s, but then turned back again, the look away time was not included in the orientation time. When the trial ended, or the infant looked away for more than 3 seconds, the side light was extinguished and the center light began to blink until the infant reoriented towards the center. At that point, one of the side lights was randomly chosen to start blinking, initiating another trial. In order to prevent any influence over the infants' looking time, both the researcher and the infant's caregiver wore sound-attenuating Peltor

headphones and listened to music so that they were unaware of the stimuli played during trials.

Design. Each experiment was split into two phases: a familiarization phase and a test phase. The familiarization phase consisted of four passages spoken by two speakers from one of the two dialects – for example, four passages of American English. Half the infants were familiarized with American English sentences and the other half with Australian English sentences. In order to move on to the test phase, infants had to listen to each passage for a total of at least 20 seconds, for a total minimum familiarization time of 80 s.

The test phase consisted of 8 test trials – 4 unheard passages of each dialect spoken by 2 new speakers per dialect. The order of presentation of the 8 test trials was randomized for each infant. The average looking times to the familiar and novel languages in the test phase was calculated for each infant and compared statistically.

4.2.2 Results and discussion

Mean orientation times to the familiar and novel dialect trials in the test phase were calculated for each infant. A two-factor repeated-measures ANOVA with Familiarization Condition as a between-subjects variable (2 levels, American vs. Australian English) and

Test Dialect as the within-subjects variable (2 levels, novel vs. familiar) was used to analyze the results for each age group.[7]

Fourteen of the 22 5-month-old infants had a longer orientation time to the novel dialect. Across all infants, the average orientation times were 9.9 s (SD = 3.7 s) for the novel dialect and 9.3 s (SD = 3.6 s) for the familiarized language, as shown in Figure 4.1. There was no significant main effect of Familiarization Condition ($F(1,20) = 0.674$; $p = 0.421$) or Test Dialect ($F(1,20) = 0.801$; $p = 0.381$), and there was no significant interaction of Familiarization and Test Dialect ($F(1,20) = 0.417$; $p = 0.526$).

Seven of the 14 7-month-old infants had a longer orientation time to the novel dialect. Across all infants, the average orientation times were 8.5 s (SD = 4.4 s) for the novel dialect and 8.5 s (SD = 3.4 s) for the familiarized dialect. There was no significant main effect of Familiarization Condition ($F(1,12) = 0.002$; $p = 0.965$) or Test Dialect ($F(1,12) = 0.213$; $p = 0.653$), and there was no significant interaction of Familiarization and Test Dialect ($F(1,12) = 0.801$; $p = 0.388$).

Six of the 8 9-month-old infants had a longer orientation time to the novel dialect. Across all infants, the average orientation times were 8.4 s (SD = 2.8 s) for the novel dialect and 7.2 s (SD = 1.7 s) for the familiarized dialect. There was no significant main effect of Familiarization Condition ($F(1,6) = 0.482$; $p = 0.405$) or Test Dialect ($F(1,6) =$

[7] An additional ANOVA was conducted with age as a between-subjects factor, but found no differences between age groups.

1.413; p = 0.280), and there was no significant interaction of Familiarization and Test Dialect ($F(1,6) = 1.560$; $p = 0.258$).[8]

An additional RM-ANOVA was conducted on the combined ages, but also found no significant main effects (Familiarization: $F(1,42) = 1.154$; $p = 0.289$; Test Dialect: $F(1,42) = 1.327$; $p = 0.256$; Interaction: $F(1,42) = 0.244$; $p = 0.624$). Thus, infants' orientation times to the novel dialect did not differ from those to the familiarized dialect for any age group, indicating that American English-learning infants between 5- and 9-months were unable to discriminate between their native dialect and a non-native dialect, Australian English.

Previous research on language discrimination ability in general suggests that this ability increases with age (Nazzi et al., 2000). Older infants should be able to discriminate language pairs that younger infants could not, and indeed, results from Chapter 3 confirm this. American English-learning 5-month-olds could not discriminate between American English and German, but 7-month-olds could. However, previous research on dialect discrimination, specifically research on the ability to discriminate between American and Australian English (Diehl et al., 2006; Kitamura et al., 2006; 2006a), suggests that the ability to discriminate between dialects of the native language declines with age. Specifically, American English 6-month-olds can discriminate between American and Australian English, but 8-month-olds cannot.

[8] Although there were no significant main effects or significant trends, the data appear to be heading in the right direction. Six of the 8 infants did look longer at the novel language than the familiar language. It is possible that this condition could become significant with additional subjects. Testing in this condition was stopped because of the high attrition rate (50%) and because of competition with other studies in the lab (see Thatte's dissertation).

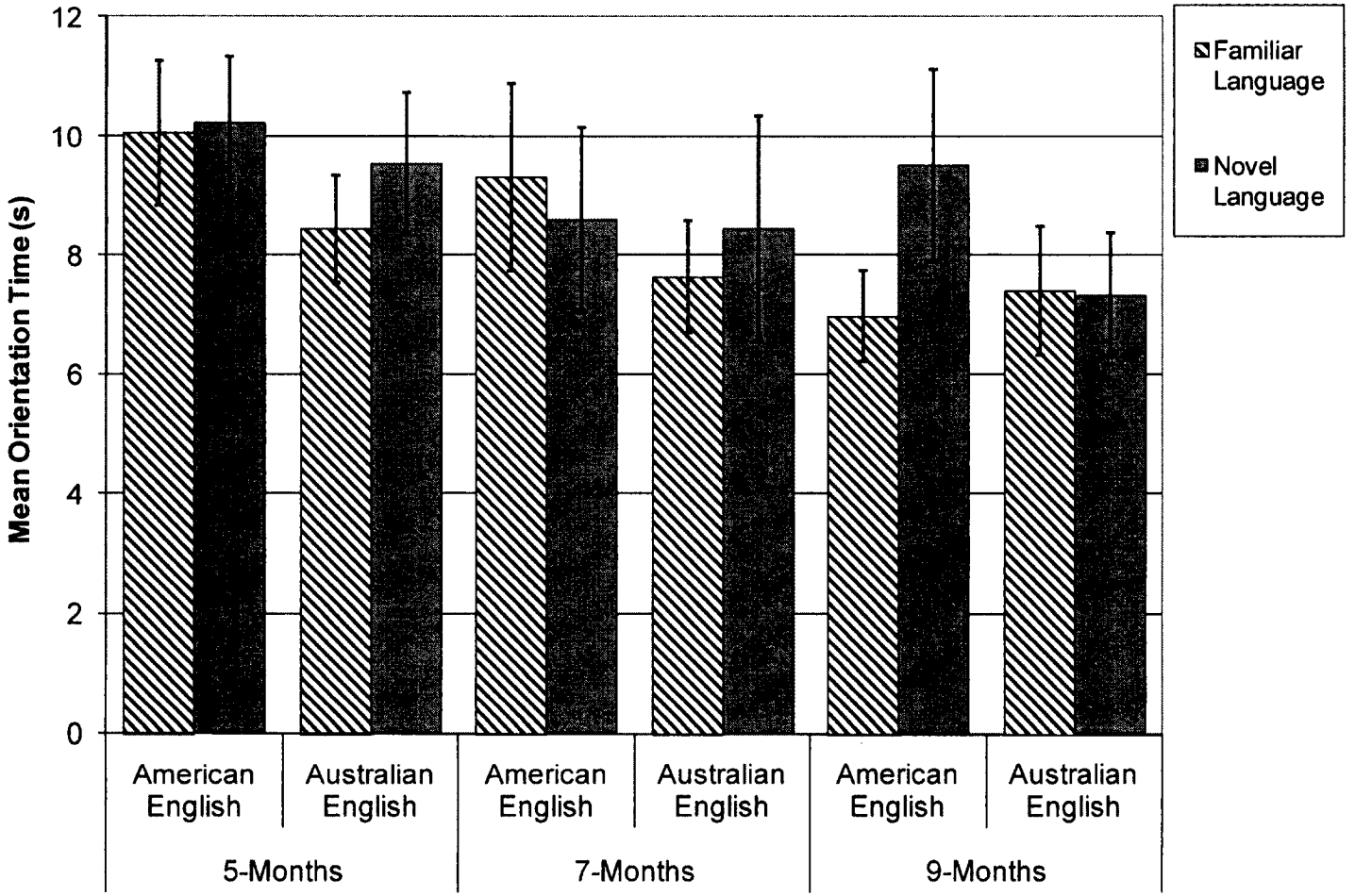

Figure 4.1. Mean orientation times (and standard error bars) to the test languages, separated by familiarization language and age group. Twenty-two 5-month-olds, fourteen 7-month-olds and eight 9-month-olds were tested. The tested languages are American and Australian English.

In the current experiment, neither pattern of development was observed. At no age between 5- and 9-months was discrimination between dialects found. This is in direct conflict with Kitamura et al. (2006), where American English-learning 6-month-olds were shown to discriminate between American and Australian English. There are methodological differences between these studies that could explain these conflicting results. Kitamura et al. (2006) tested preference, while infants in this experiment were tested on a more direct discrimination paradigm. Yet, this fact is at first damning, as any differences based on that account should show the reverse pattern. Infants that can discriminate may not show a preference, but infants that show a preference should be able to discriminate.

However, there were also differences in the complexity of the stimuli. While both studies used multiple speakers (four from each dialect), Kitamura et al.'s (2006) stimuli consisted of only 5 unique sentences, while the stimuli used in the current study consisted of 40 unique sentences. Thus, while infants had to abstract over multiple speakers in both studies, they had to abstract over a greater amount of linguistic variation in the current experiment than in Kitamura et al. (2006). Butler et al. (2011) argued that the more repetitive stimuli used in Kitamura et al. (2006) allowed infants, especially the older infants, to develop a more abstract representation of the language, causing them to ignore the phonetic variations between the two dialects. They suggest this is the reason why Kitamura et al. (2006) found a decline in dialect discrimination ability with age while they found no similar decline in British English-learning 5- and 7-month-olds' ability to discriminate between different dialects of British English.

There was also a difference in speaking style. The stimuli used in the current experiment were recorded as rather emotionless, adult-directed speech, while the stimuli in Kitamura et al. (2006) was infant-directed speech, which has been shown in many studies to be more preferable to infants than adult-directed speech (Cooper and Aslin, 1990; Cooper et al., 1997). Thus, the more complex and less interesting stimuli used in the current experiment may have hindered infants' ability to discriminate between the two dialects. Infants may not be able to discriminate between this type of dialectical stimuli until after 9-months of age.

Still, the stimuli and methodologies used in the current experiment are similar to Nazzi et al. (2000) and Butler et al. (2011), and both those studies found discrimination between dialects by 5-months. Recall also that Nazzi et al. (2000) found that American English-learning 5-month-olds could discriminate between prosodically similar languages, British English and Dutch, as well, and that Bosch and Sebastián-Gallés (1997) showed that Spanish- and Catalan-learning 4-month-olds could also discriminate between prosodically similar languages, Spanish and Catalan. However, neither the current experiment nor the experiments in Chapter 3 were able to find evidence of discrimination at 5-months. It would be reassuring to replicate the finding that American English 5-month-olds can discriminate some language pairs.

The following experiments test American English-learning 5-month-olds on their ability to discriminate between American English and Dutch, British English and Dutch, and American English and Japanese.

4.3 Experiment 4.2 - American English and Dutch

Nazzi et al. (2000) showed that American English-learning 5-month-olds could discriminate between British English and Dutch. Therefore, it was expected that American English-learning 5-month-olds should also be able to discriminate between American English and Dutch.

4.3.1 Methods

Participants. Eighteen American 5-month-olds (9 males and 9 females) from monolingual English-speaking homes participated. The mean age was 150 days (range = 135 to 170 days). An additional 4 infants were tested but their data were not included because they failed to complete the experiment due to fussiness and crying (2) or lack of interest (2).

Stimuli. The stimuli consisted of the 8 American English passages used in Experiment 4.1, and 8 Dutch passages, taken from Nazzi et al. (2000). Passages were normalized for intensity in Praat to 80 dB. The acoustic properties of the stimuli are given in Table 4.2, including duration and mean f0, and f0 range. Mean number of syllables is not provided because there was no access to a written transcription of the Dutch passages. However, sentences were originally chosen in Nazzi et al. (2000) to match in number of syllables.

	American English	Dutch
Average Duration	17.2 s (1.0)	16.7 s (0.3)
Mean f0	215 Hz (21)	221 Hz (18)
Mean f0 Range	466 Hz (75)	406 Hz (120)

Table 4.2. Acoustic properties of the American English and Dutch passages.

Procedure and Design. The procedure and design used in this experiment were identical to Experiment 4.1.

4.3.2 Results and discussion

Mean orientation times to the familiar and novel language trials presented in the test phase were calculated for each infant. Eleven of the 18 infants had a longer orientation time to the novel language. Across all infants, the average orientation times were 9.4 s (SD = 3.3 s) for the novel language and 8.8 s (SD = 3.3 s) for the familiarized language, as shown in Figure 4.2. A two-factor repeated-measures ANOVA with Familiarization Condition (2 levels, English vs. Dutch) as a between-subjects variable and Test Language (2 levels, novel vs. familiar) as the within-subjects variable showed no significant main effect of Test Language ($F(1,16) = 0.404$; $p = 0.534$), or Familiarization Condition ($F(1,16) = 1.034$; $p = 0.324$), and no interaction between Familiarization and Test Language($F(1,16) = 0.080$; $p = 0.781$). Thus, infants' orientation times to the novel dialect did not differ from those to the familiarized dialect indicating that American English-learning 5-month-olds were unable to discriminate between their native language and a non-native language, Dutch.

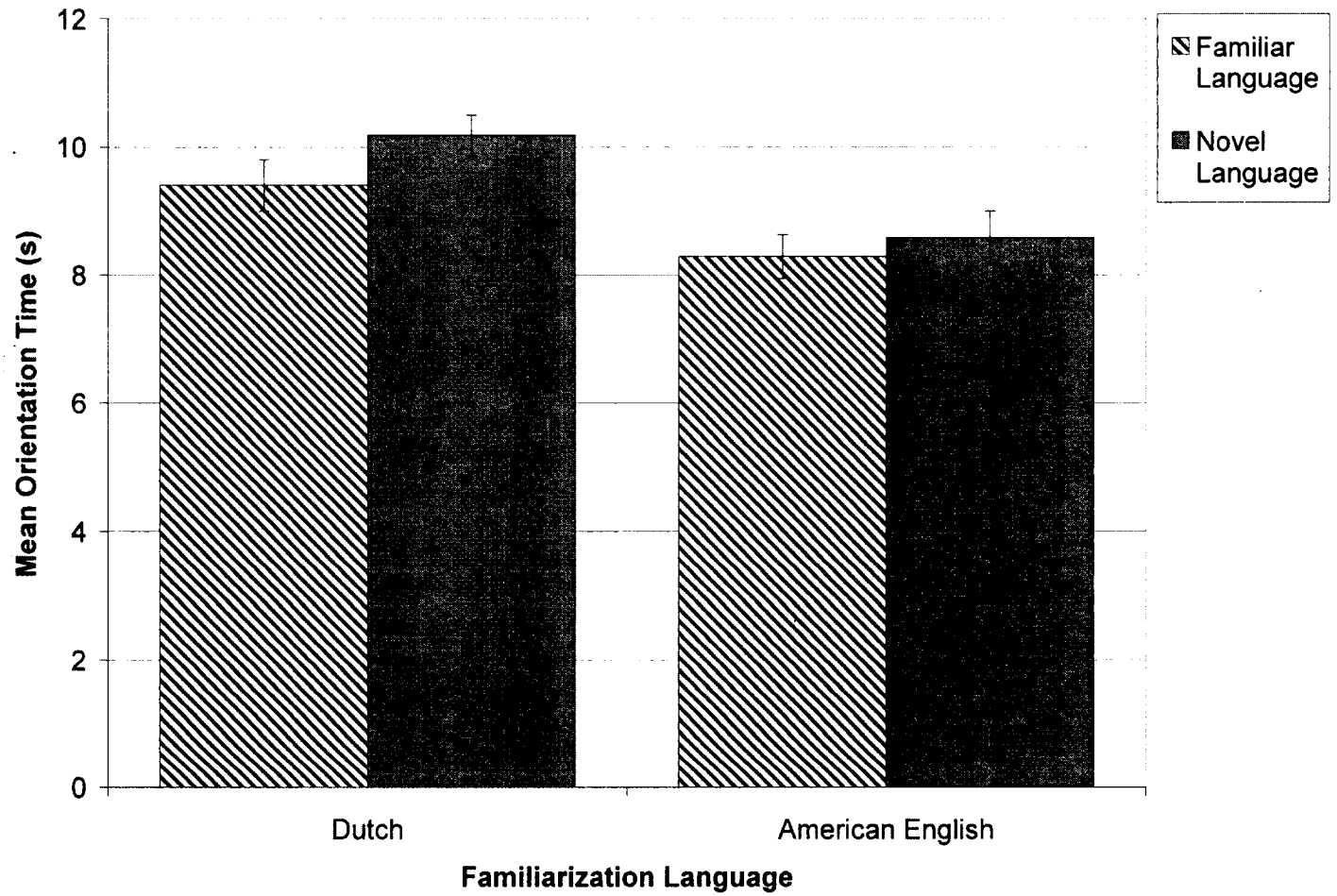

Figure 4.2. Mean orientation times (and standard error bars) of American English-learning 5-month-olds to the test languages, separated by familiarization language. The tested languages are American English and Dutch.

As before, these results are unexpected because Nazzi et al. (2000) showed that American English-learning 5-month-olds could discriminate between British English and Dutch. However, they also showed that American English-learning 5-month-olds could discriminate between American and British English. Thus, while such a fact would be surprising, it is possible that in their study, American English-learning 5-month-olds were treating both American English and Dutch as their native language and British English as non-native. Or at least, American English and Dutch were judged to be the same language, and British English as a different language. If so, then infants should fail to discriminate between American English and Dutch, as they do in the current experiment.

Similarly, Nazzi et al. (2000) found American English-learning 5-month-olds could not discriminate between German and Dutch. As shown in chapter 3, American English-learning 5-month-olds also fail to discriminate between American English and German. Again, if to American English-learning 5-month-olds, Dutch is indistinguishable from American English, then they should show the same discrimination pattern on Dutch and German as they do on American English and German. Indeed, 5-month-olds fail to discriminate both pairs. Therefore, under this interpretation, British English is the oddball. American English-learning 5-month-olds view it as distinct while American English, German and Dutch are viewed as the same. In any case, because null results were once again obtained, evidence that 5-month-olds can discriminate languages has not yet been replicated.

4.4 Experiment 4.3 – British Englishand Dutch

In this experiment, a more direct replication of Nazzi et al. (2000) was attempted. American English-learning 5-month-olds were tested on their ability to discriminate between British English and Dutch. Like in the previous experiments, the testing procedure and apparatus was identical to that used in Nazzi et al. (2000). In addition, the same stimuli used in Nazzi et al. (2000) were used in this experiment.

4.4.1 Methods

Participants. Eighteen American 5-month-olds (10 males and 8 females) from monolingual English-speaking homes participated. The mean age was 151 days (range = 133 to 168 days). An additional 10 infants were tested but their data were not included because they failed to complete the experiment due to fussiness and crying (5) or lack of interest (4), or because of experimental error (1).

Stimuli. The stimuli consisted of 8 British English passages and 8 Dutch passages, taken from Nazzi et al. (2000). Passages were normalized for intensity in Praat to 80 dB. The acoustic properties of the stimuli are given in Table 4.3, including duration and mean f0, and f0 range. Mean number of syllables is not provided because there was no access to a written transcription of the Dutch passages. However, sentences were originally chosen in Nazzi et al. (2000) to match in number of syllables.

	British English	Dutch
Average Duration	16.5 s (0.2)	16.7 s (0.3)
Mean f0	238 Hz (15)	221 Hz (18)
Mean f0 Range	415 Hz (92)	406 Hz (120)

Table 4.3. Acoustic properties of the British English and Dutch passages.

Procedure and Design. The procedure and design used in this experiment were identical to Experiment 4.1.

4.4.2 Results and discussion

Mean orientation times to the familiar and novel language trials presented in the test phase were calculated for each infant. Ten of the 18 infants had a longer orientation time to the novel language. Across all infants, the average orientation times were 8.6 s (SD = 3.5 s) for the novel language and 9.0 s (SD = 3.0 s) for the familiarized language, as shown in Figure 4.3. A two-factor repeated-measures ANOVA with Familiarization Condition (2 levels, English vs. Dutch) as a between-subjects variable and Test Language (2 levels, novel vs. familiar) as the within-subjects variable showed no significant main effect of Test Language ($F(1,16) = 1.832$; $p = 0.195$), or Familiarization Condition ($F(1,16) = 1.864$; $p = 0.191$). However, there was a significant interaction between Familiarization and Test Language ($F(1,16) = 9.966$; $p = 0.006$). Surprisingly, regardless of the language they were familiarized to, infants oriented longer to British English than to Dutch. Fifteen of the 18 infants had a longer orientation time to British English. Across all infants, the average orientation times were 8.3 s (SD = 3.5 s) for Dutch and 9.3 s (SD = 3.0 s) for British English.

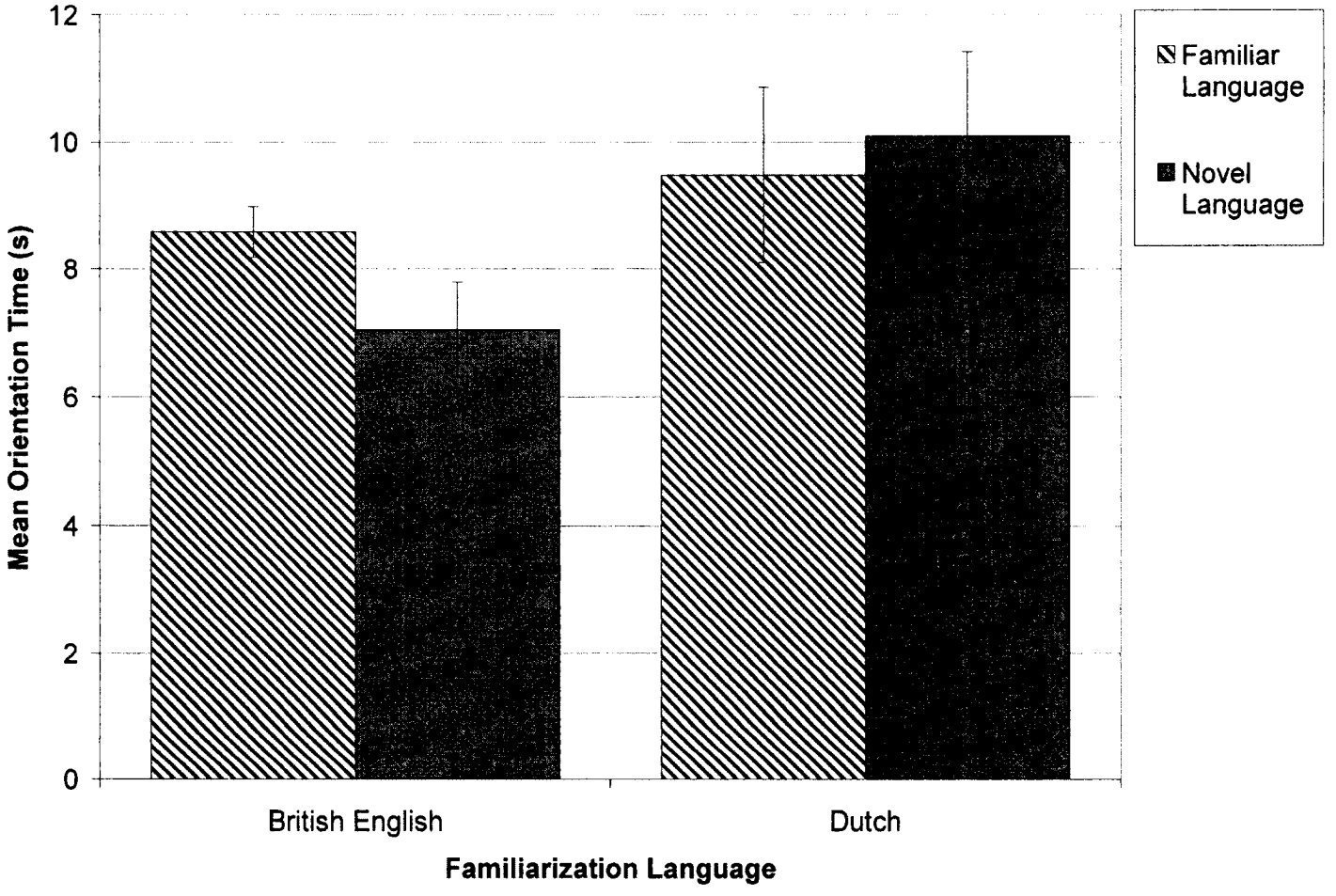

Figure 4.3. Mean orientation times (and standard error bars) of American English-learning 5-month-olds to the test languages, separated by familiarization language. The tested languages are British English and Dutch.

Contrary to previous HPP discrimination experiments in the literature, infants in this experiment showed neither the expected novelty preference nor a familiarity preference. Rather, infants showed a preference for British English, regardless of the language they were familiarized to. Regardless, these results contrast with those from Nazzi et al. (2000), where infants showed a more typical novelty preference. Thus, on the most basic level, their results failed to replicate. However, in order to prefer one language over the other, infants must be able to discriminate between them. Therefore, it can be said conclusively that American English-learning 5-month-olds were able to discriminate between British English and Dutch, which is what Nazzi et al. (2000) found. It is unclear, though, why they observed a novelty preference, but a preference for British English was seen in this experiment, even though the stimulus set is the same. The next experiment tests discrimination on a prosodically dissimilar language pair, American English and Japanese.

4.5 Experiment 4.4 – American English and Japanese

This experiment tests American English-learning 5-month-olds on their ability to discriminate between prosodically dissimilar languages American English and Japanese. English is a lexical stress language that marks stressed syllables with a variety of semantically meaningful pitch contours. Japanese is a lexical pitch accent language that marks the accented syllable with a single pitch contour, a falling contour. English is intonationally considered a head-marking language, while Japanese is considered both a

head- and edge-marking language (Jun, 2005). Rhythmically, English is considered a stress-timed language, while Japanese is considered to be a mora-timed language.

Although this specific language pair has never before been tested in a language discrimination or preference study (British English and Japanese have been tested (Christophe and Morton, 1998; Nazzi et al., 1998; Nazzi et al., 2000) as have Australian English and Japanese (Hayashi et al., 2001; Sundara et al., 2008)), according to the rhythm hypothesis, infants should be capable of discriminating between languages from different rhythm classes at birth (Nazzi et al., 1998). Indeed, Nazzi et al. (2000) showed that American English-learning 5-month-olds could discriminate between British English and Japanese.

4.5.1 Methods

Participants. Eighteen American 5-month-olds (9 males and 9 females) from monolingual English-speaking homes participated. The mean age was 157 days (range = 135 to 168 days). An additional 3 infants were tested but their data were not included because they failed to complete the experiment due to fussiness and crying (1), lack of interest (1), or because they wet themselves (1).

Stimuli. The stimuli consisted of the 8 American-English passages used in Experiment 4.1, and 8 Japanese passages, taken from Nazzi *et al.* (2000). Passages were normalized for intensity in Praat to 80 dB. The acoustic properties of the stimuli are given in Table 4.4, including duration, mean f0, and f0 range. Mean number of syllables

is not provided because there was no access to a written transcription of the Japanese passages. However, sentences were originally chosen in Nazzi et al. (2000) to match in number of syllables.

	American English	Japanese
Average Duration	17.2 s (1.0)	18.5 s (0.1)
Mean f0	215 Hz (21)	239 Hz (18)
Mean f0 Range	466 Hz (75)	364 Hz (115)

Table 4.4. Acoustic properties of the American English and Japanese passages.

Procedure and Design. The procedure and design used in this experiment were identical to Experiment 4.1.

4.5.2 Results and discussion

Mean orientation times to the familiar and novel language trials presented in the test phase were calculated for each infant. Ten of the 18 infants had a longer orientation time to the novel language. Across all infants, the average orientation times were 8.9 s (SD = 2.4 s) for the novel language and 8.7 s (SD = 3.3 s) for the familiarized language, as shown in Figure 4.4. A two-factor repeated-measures ANOVA with Familiarization Condition (2 levels, English vs. Japanese) as a between-subjects variable and Test Language (2 levels, novel vs. familiar) as the within-subjects variable showed no significant main effect of Test Language ($F(1,16) = 0.123$; $p = 0.730$), or Familiarization Condition ($F(1,16) = 0.241$; $p = 0.630$). However, the interaction between Familiarization and Test Language trended towards significance ($F(1,16) = 4.314$; $p =$

0.054). Regardless of the language they were familiarized to, infants oriented longer to American English than to Japanese. Thirteen of the 18 infants had a longer orientation time to American English. Across all infants, the average orientation times were 8.1 s (SD = 2.8 s) for Japanese and 9.5 s (SD = 2.8 s) for American English.

Again, infants failed to show the expected novelty effect or a familiarity effect. Rather, they showed a preference for their native language, American English, regardless of the language they were familiarized to. Many previous studies on language preference have shown that infants prefer their native language over a non-native language (Moon et al., 1993; Hayashi et al., 2001; Sundara et al., 2008; Byers-Heinlein et al., 2010). The results from Experiment 4.3 and the current experiment fit with that pattern, as infants prefer the English dialect over the non-English language in both cases.[9]

Both Hayashi et al. (2001) and Sundara et al. (2008) tested infants between 4- and 14-months, but found no language preference in the younger infants. Infants in these studies showed no native language preference until 7-months or older. Thus, it is odd that 5-month-olds are showing a language preference in the current experiments. However, infants younger than 4-months were found to show a native language preference in Moon et al. (1993) and Byers-Heinlein et al. (2010). These latter studies tested 2-month-olds and newborns, respectively, using the High Amplitude Sucking procedure (HAS), while the former studies used HPP. Thus, one possible explanation for the conflicting results might be the difference in testing methodology.

[9] However, if as suggested, infants are treating American English and Dutch as the same language and British English as distinct, then the British preference observed in the current experiment is recast as a non-native preference, in contrast with the previous preference research.

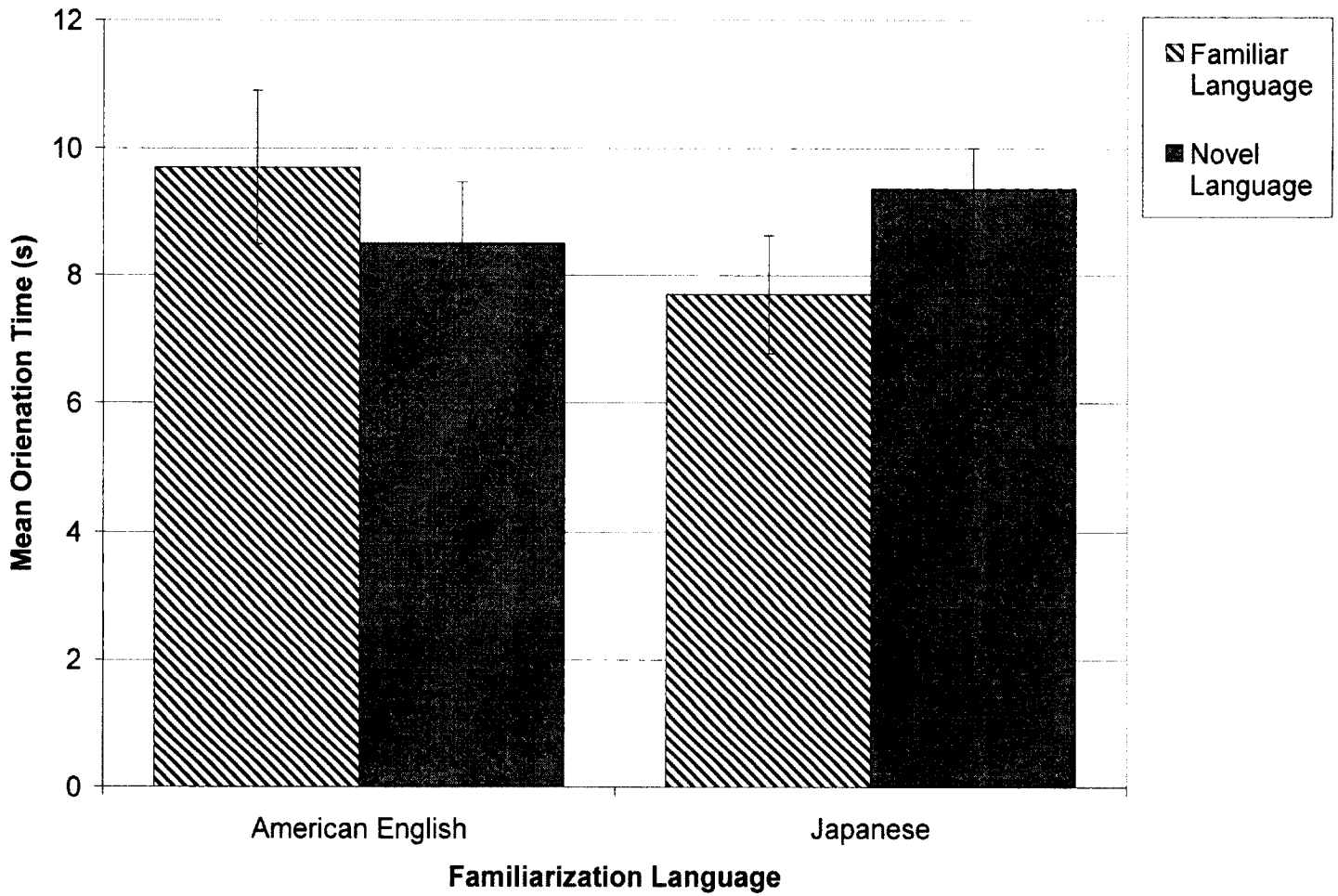

Figure 4.4. Mean orientation times (and standard error bars) of American English-learning 5-month-olds to the test languages, separated by familiarization language. The tested languages are American English and Japanese.

However, another difference between the studies was their length. The HPP preference studies (Hayashi et al., 2001; Sundara et al., 2008) were short, on the order of 5 minutes in length. The HAS studies (Moon et al., 1993; Byers-Heinlein et al., 2010), on the other hand, were longer, between 10-18 minutes, and, in Moon et al. (1993), no preference was observed until the final 6 minutes of the experiment (or, after infants had been listening for at least 12 minutes). The longer experiments may have bored or exhausted infants to the point where they prefer to listen to something familiar – their native language over a non-native language. This would explain why Moon et al. (1993) found no preference until the final third of their experiment.

The current experiments, like Hayashi et al. (2001) and Sundara et al. (2008), were also only about 5 minutes in length. However, in their studies, infants heard two languages intermixed with one another, while in the current experiments, infants were exposed to only one language for an average of 140 s before hearing a second. Thus, it is possible infants in the current studies, like those in Moon et al. (1993), were sufficiently bored by the test phase that they showed a preference for their native language.

As in Experiment 4.3, because 5-month-olds preferred one of the languages over the other, it can be concluded that they were able to discriminate between them. It is still unclear, though, why infants are showing a specific language preference in Experiments 4.3 and 4.4 rather than the expected novelty preference, and null results in other experiments when discrimination is expected. Because American English and Japanese are considered to be quite dissimilar in their prosodic and phonotactic properties, it is thought they should be distinguishable at birth, let alone at 5-months of age. Therefore,

the difficulty found in replicating or extending the results from Nazzi et al. (2000) does not appear to be due to the prosodically similar nature of the tested languages.

Perhaps the difficulty is more basic, such as experimenter errors? This is unlikely because participants in these experiments were tested by multiple experimenters, including both graduate students and undergraduate researchers. In these experiments, all infant participants showed similar behaviors – they had similar familiarization times and test language looking times. No single experimenter was responsible for testing the infants that fell towards the outlying ranges of these measures. These experimenters also conducted the experiments in Chapter 3, two of which showed the expected novelty preference, indicating that at least some of the HPP discrimination experiments can produce expected results. Similarly, these experimenters all test infants in HPP for other studies unrelated to this project, including word segmentation experiments similar to Jusczyk et al. (1999), which have produced results fitting with previously established patterns (Kim, p.c.).

The American English-learning infants tested in Nazzi et al. (2000) were tested at John Hopkins University, while the American English-learning infants tested in the current experiments were tested at UCLA. Perhaps the difference lies in the participating infants? This, too, seems unlikely. Although they have been treated thus far as the same dialect, Southern Californian English could be distinct from Baltimorean English (certainly, there are segmental differences, such as the layout of the vowel space), and that difference could impact discrimination ability. In as much as this has been shown to be true, though, language background matters most when discriminating between non-

native languages or dialects, essentially a harder task (Ramus and Mehler, 1999; Nazzi et al., 2000; Barkat and Vasilescu, 2001; Butler et al., 2011). All the experiments here except for Experiment 4.3 test Southern Californian English-learning infants on Southern Californian English and some non-native language or dialect, which should be easier than in Nazzi et al. (2000) where Baltimorean English-learning infants are tested on British English and a non-native language. In the case of Experiment 4.3, a direct replication of Nazzi et al. (2000), like their Baltimorean English-learning infants, Southern Californian English-learning infants are tested on two non-native languages, British English and Dutch. Yet, normal discrimination behavior is observed in the Nazzi et al. (2000) experiments, while null results or atypical behavior are found here.

Finally, perhaps the difficulty in replicating or extending the results from Nazzi et al. (2000) doesn't have to do with the tested languages, experimenters or participants, but rather with the testing methodology itself? Nazzi et al. (2000) and Bosch and Sebastián-Gallés (2001) were the first researchers to use HPP in a familiarization-preference design to test language discrimination, and they both found a novelty preference. However, Butler et al. (2011) also used HPP in a familiarization-preference design to test 5- and 7-month-old South-West British English-learning infants on their ability to discriminate between their native dialect and a non-native dialect, Welsh English. For 5-month-olds, they found infants could discriminate, but showed a familiarity preference. Seven-month-olds, on the other hand, showed a novelty preference.

Houston-Price and Nakai (2004), in reviewing novelty vs. familiarity effects in infant research, point out three factors that can influence the direction of this effect – the

length of the experiment, infant age, and the complexity or salience of experimental stimuli (Hunter and Ames, 1988). Butler et al. (2011) used infants of the same age and an identical procedure as Nazzi et al. (2000). However, they used infant-directed speech while Nazzi et al. (2000) used adult-directed speech. Butler et al. (2011) suggest that the more salient stimuli in their experiment prevented infants from habituating in the familiarization phase, leading to a familiarity effect over a novelty effect. Older infants, on the other hand, habituate faster than younger infants, so the more salient stimuli did not prevent 7-month-olds from habituating during the familiarization phase.

While this suggestion provides a reasonable explanation for their specific data, it highlights a general weakness with this experimental setup. Namely, a novelty preference is expected to result in the familiarization-preference design because infants are treated as habituated during the familiarization phase. However, recall that in a familiarization-preference design, infants are exposed to a fixed duration of stimuli. Individual boredom or habituation is not measured. Thus, after completing the familiarization phase, some infants may be habituated, while some may not.

As implied by Butler et al. (2011), this may lead some infants to show a novelty effect while others show a familiarity effect, wiping out any group effect. Thus, in the case of Experiment 4.3, some American English-learning 5-month-olds may be showing a familiarity effect, while others show a novelty effect when discriminating between American English and Dutch. It may indeed be the case that 5-month-olds can discriminate between American English and Dutch, but the inconsistent habituation of the HPP familiarization-preference setup is obscuring this fact.

It is possible the specific language preference could result from a fixed familiarization time used in the familiarization-preference design, as well. For example, a lack of true habituation in a stressful experiment or over-habituation in a long experiment might cause infants to fall back on their cumulative language experience, causing them to orient towards their native language. If uncontrolled habituation is playing a role in American English-learning 5-month-olds preference for British English over Dutch in Experiment 4.3 and for American English over Japanese in Experiment 4.4, then this preference should disappear in an experiment where infant habituation is more directly controlled.

4.6 Experiment 4.5 – British English and Dutch in visual habituation

Results on American English-learning 5-month-olds' discrimination of British English and Dutch using HPP in a familiarization-preference design are inconsistent. Nazzi et al. (2000) found that infants showed a novelty effect, while in Experiment 4.3, above, we found that infants showed a preference for British English, regardless of the familiarization language. These differing results may be due to the lack of controlled habituation in the design. In this experiment, American English-learning 5-month-olds' ability to discriminate British English and Dutch was tested in a Visual Habituation Paradigm (VisHab) with an infant-controlled habituation phase.

In infant-controlled habituation experiments, infants are exposed to a number of trials and their attention to each is measured. Over the course of many trials, infant

attention will naturally decline as they become bored. Only when attention, averaged over some set of sequential trials, drops below a certain threshold do infants move out of the habituation phase and into the test phase of the experiment. The time this takes can differ from infant to infant. In the testing phase, there is a strong expectation for infants to show a novelty effect because infants will be bored of the habituated stimuli.

In their report, Nazzi et al (2000) criticize the VisHab paradigms used in earlier discrimination studies (Mehler et al., 1988; Christophe and Morton, 1998) as being less powerful in comparison to their HPP familiarization-preference design because the previous VisHab studies were based on a between-subjects design, while their HPP study used a within-subjects design. To address this criticism, in the current experiment, infants were habituated to a single language and then tested on unheard passages from both languages, making it a within-subjects design.

4.6.1 Methods

Participants. Eighteen American 5-month-olds (4 males and 14 females) from monolingual English-speaking homes participated. The mean age was 156 days (range = 138 to 168 days). An additional 10 infants were tested but their data were not included because they failed to complete the experiment due to fussiness and crying (3), because they maxed out on the test trials (5), because they failed to dishabituate to the post-test stimuli (1), or because they were deemed to show false habituation (their familiar

language looking time was larger than the novel language by over 2 standard deviations; 1)

Stimuli. The stimuli consisted of the 8 American-English passages and 8 Dutch passages, used in Experiment 4.3. In addition, a passage of adult-directed Spanish, spoken by a female speaker was used for the post-test trial. This passage was taken from Ward (2011).

Procedure. Infants were tested in the Visual Habituation Paradigm. Each infant was tested individually while seated on her parent's lap in front of a television screen. A loudspeaker and camera were hidden below the television screen. The camera was connected to a monitor, and the television display and speakers were controlled by a computer, both located outside the booth.

At the beginning of each trial, the infant's attention was attracted to the television screen by displaying a zooming bull's eye and playing a recording of an infant giggling. When the infant oriented towards the screen, the image was switched to a black and white checkerboard an auditory stimulus began to play. The stimulus continued to play until the recording ended (max duration = 17 s), or until the infant failed to maintain orientation towards the TV screen for 2 consecutive seconds. A researcher seated at the computer terminal outside the room recorded the duration of the infant's eye gazes towards the screen. If the infant looked away for less than 2 s, but then turned back again, the look-away time was not included in the orientation time. When the trial ended, the checkerboard was removed from the screen, and a new trial was initiated. In order to prevent any influence over the infants' looking time, both the researcher and the infant's

caregiver wore sound-attenuating Peltor headphones and listened to music so that they were unaware of the stimuli played during trials. Stimulus presentation and looking time measurement was controlled by HABIT X (Cohen et al., 2000).

Design. Each experiment was split into two phases: a habituation phase and a test phase. The stimuli used for habituation consisted of six passages spoken by three speakers from one of the two languages – for example, six passages of British English. The order of presentation for the habituation trials was counterbalanced across subjects. Half of the infants were habituated to British English and half to Dutch. In order to move onto the test phase, infants' orientation time averaged over the final three habituation trials had to decline by 40% from the orientation time averaged over the initial three trials. Thus, the length of the habituation phase was infant-controlled, and lasted for a minimum of 4 trials (average = 10).

The test phase consisted of 3 trials – a test trial, a control trial and a post-test trial. The control trial was an unheard passage from an unheard speaker of the habituation language, and the test trial was a passage from the novel language. The post-test trial was a passage of Spanish, which was meant to serve as a prosodically distinct, easily distinguishable language. The order of the test and control trials was counterbalanced across subjects. The post-test trial was always presented last. The average looking times to the different test phase trials was calculated for each infant and compared statistically.

4.6.2 Results and discussion

Eleven of the 16 infants had a longer orientation time to the novel language than to the familiar language. Across all infants, the average orientation time was 9.2 s (SD = 5.6 s) for the novel language, 5.8 s (SD = 2.6 s) for the familiar language, and 12.1 s (SD = 6.2 s) for the post-test/Spanish trial, as shown in Figure 4.5. Orientation times in three test phase trials were compared to the average orientation time on the final three habituation trials using a repeated-measures ANOVA with Trial (4 levels: Habituation, Control, Test, Post-test) as a within-subjects variable and Habituation Language (2 levels: British vs. Dutch) as a between-subjects variable. There was a significant main effect of Trial type ($F(3,48) = 7.897$; $p < 0.001$), but there was no effect of Habituation Language ($F(1,16) = 0.439$; $p = 0.517$) nor any interaction ($F(3,48) = 1.965$; $p = 0.132$).

Planned paired t-tests were used as post-hoc analysis to examine differences between trial types. There was no significant difference between the final three habituation trials and the control trial ($t(17) = 0.859$; $p = 0.402$). Thus, infants treated new passages in the same language as equivalent to the habituation stimuli. There was a significant difference between the final three habituation trials and test trials ($t(17) = 2.330$; $p = 0.032$). Infants dishabituated and listened longer towards the novel language, indicating that infants were able to successfully discriminate between British English and Dutch. There was no difference between the test and post-test trials ($t(17) = 1.271$; $p =$

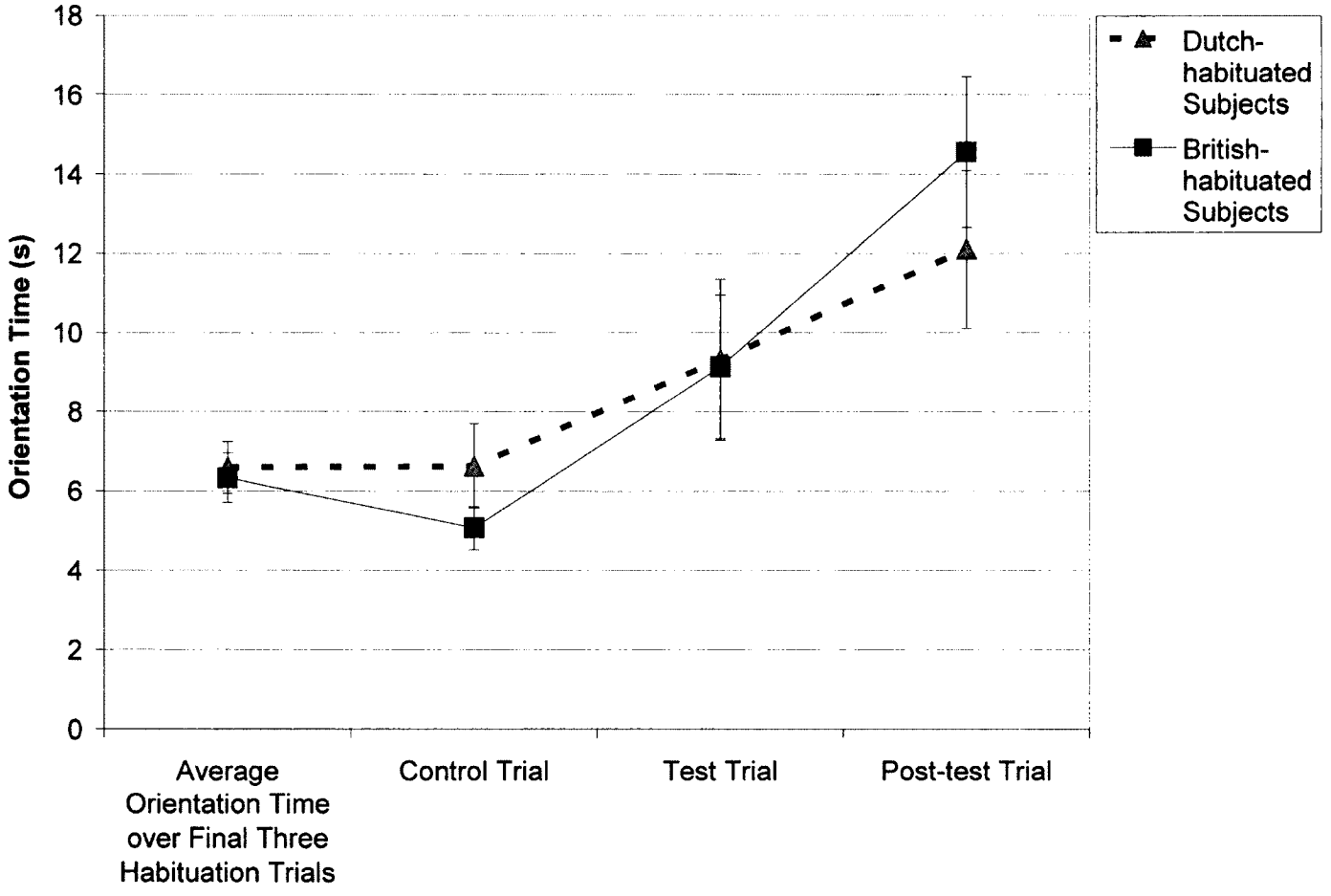

Figure 4.5. Orientation times (and standard error bars) of American English-learning 5-month-olds to the final three habituation trials and the three test phase trials. Results for all currently tested subjects, and for the subset habituated to British English are given. The tested languages are British English and Dutch. The post-test language is Spanish.

0.221), but there was a difference between the control and post-test trials ($t(17) = 4.387$; $p < 0.001$), suggesting that infants were also able to discriminate between British English/Dutch and Spanish.

These results confirm that American English-learning 5-month-olds can indeed discriminate between British English and Dutch. As expected, when habituation is more carefully controlled, infants show the expected novelty effect. In contrast, experiments using HPP with a familiarization-preference design to test American English-learning 5-month-olds on this language pair show differing results. Nazzi et al. (2000) showed a novelty effect, while a preference for British English sentences was seen in experiment 3 in this chapter.

We can compare the two experiments testing 5-month-olds' discrimination of British English and Dutch to better understand this discrepancy. When tested with HPP, the average listening time during the familiarization phase was 142 s (SD = 22.5). Using visual habituation, the average total habituation time was 123 s (SD = 83.6), which was not significantly different ($t(34) = 0.926$; $p = 0.361$) from the habituation time using HPP. Crucially, the range of habituation times in the visual habituation study was far greater. In the HPP study, the infant who moved out of the familiarization phase fastest listened for 114.5 s, while the infant who moved out slowest listened for 186.1 s. In contrast, of the 18 infants tested in the VisHab study, ten infants habituated faster than 114.5 s (as fast as 35.6 s) and five took longer than 186.1 s (as long as 338.1 s). Thus, with familiarization times restricted between 118 and 186 s, only a small subset of the infants would have been properly habituated. Thus, it is likely that, in Experiment 4.3, the HPP

study, and in Nazzi et al. (2000), some infants had not yet habituated when they moved out the familiarization phase, and some infants were overly habituated. Therefore, their behavior would reasonably have been more unpredictable.

4.7 General discussion

In this chapter, American English-learning infants were tested on their ability to discriminate between several language pairs. Using the same methodologies as in Chapter 3, and in Nazzi et al. (2000), 5-, 7-, and 9-month-olds were tested on their ability to discriminate between American and Australian English. No discrimination was found at any age. These results contrast with previous results indicating English-learning infants can discriminate their native dialect from a non-native dialect at this age (Nazzi et al., 2000; Kitamura et al., 2006; Butler et al., 2011).

Five-month-olds were also tested on their ability to discriminate between American English & Dutch, British English & Dutch, and American English & Japanese. Infants failed to discriminate between American English and Dutch, but were able to discriminate between British English and Dutch, and between American English and Japanese. However, for both of the latter language pairs, infants failed to show the expected novelty preference. Rather, infants showed a preference for a specific language – British English over Dutch, and American English over Japanese. These results contrast with previously published data (Nazzi et al., 2000).

The best explanation for these contrasting results seems to be that the testing method – HPP in a familiarization-preference design – is unreliable. This testing procedure has been used to test infants' language discrimination ability in few studies. The first researchers to use this method (Nazzi et al. 2000; Bosch & Sebastian-Galles, 2001) found either a novelty preference, which was interpreted as evidence of language discrimination, or no preference, which was interpreted as an inability to discriminate. However, Butler et al. (2011) found a familiarity preference, as well as a novelty preference and no preference. In the current study, possible results included a novelty preference and no preference, as well, but also a preference for one of the tested languages regardless of familiarization language. These varying possible outcomes likely result from a lack of controlled habituation by infants to the language used in the familiarization phase. When American English 5-month-olds were retested on their ability to discriminate between British English and Dutch using a different methodology that controlled for infant habituation, the expected novelty effect was observed.

Although the inability to predict the outcome of infant behavior in HPP familiarization-preference language discrimination tasks is a weakness, this methodology can still be used to learn about infants' discrimination abilities. Whenever a preference is observed – novel, familiar, or otherwise – it can be concluded that infants were able to discriminate between the tested languages. However, when no preference is observed, it is not necessarily the case that infants are unable to discriminate between the languages. Rather, infants may be capable of discrimination, but some are showing one type of preference while some show another, wiping out any group effect. These differences

might also prevent discrimination results from being easily replicated, as was the case here for American English-learning 5-month-olds and British English & Dutch. As a result, it would be best if future language discrimination studies on infants were conducted using testing methods that more directly control for infant habituation.

5. Conclusion

In this dissertation, I set out to explore adults' and infants' ability to discriminate their native language from prosodically similar non-native languages. Specifically, the question of whether (if at all) intonation plays a role in the ability to discriminate languages was investigated.

5.1 Adults' discrimination ability

Previous research indicates that both infants and adults can discriminate languages using only prosodic cues, including prosodically similar languages (Barkat, Ohala and Pellegrino, 1999; Bosch & Sebastián-Gallés, 1997). Adults have been shown to be capable of using both rhythm – the duration and timing of different segmental intervals – and intonation – the patterns of pitch contours over an utterance – to discriminate between languages. There is some evidence that adults can even use intonational differences to discriminate between prosodically similar languages or dialects (Willems, 1982; de Pijper, 1983; Ménard *et al.*, 1999). However, the relative ranking of rhythm and intonation cues had not been previously explored. These questions were investigated for adult listeners in Chapter 2. It was hypothesized that adults would be successful at distinguishing their native language from a prosodically similar non-native language using only prosodic cues. It was also hypothesized that, while listeners might be capable

of using only rhythmic information to discriminate, intonation would also play an important role in this task.

Using a language identification task, American English-speaking adults were tested on their ability to discriminate American English from a prosodically similar non-native language, German, as well as from a non-native dialect, Australian English. Subjects were presented isolated sentences and asked to identify them as either 'American English' or some 'other' language. Sentences were modified in various ways to eliminate different types of acoustic information. Sentences were:

(1) Low-pass filtered, a process which severely degrades segmental information, but preserves prosody.

(2) Re-synthesized into '?a?a?a speech' by replacing all sonorant segments with /a/ and all obstruent segments with silence (/?/), and then replicating the original pitch track. This method completely removes segmental information, but preserves rhythm and intonation.

(3) Re-synthesized into 'flat ?a?a?a speech' by replacing the original intonation with a monotone pitch. These stimuli contained only rhythmic information.

(4) Scrambled by dividing the '?a?a?a speech' sentences into smaller segments and randomly reordering them into new sentences. These stimuli contained both rhythm and intonational information, but neither was representative of American English.

The stimuli were examined acoustically and it was found that both language pairs significantly differed in their rhythm and intonation, despite their prosodic similarity. Perception experiments showed that American-English-listeners were indeed able to discriminate between both language pairs using only prosodic cues. For example, adults were able to discriminate both language pairs using low-pass filtered stimuli.

Additionally, I found that adults could use both rhythm and intonation to discriminate languages, but they weighted intonational information above rhythmic information. However, different cues were useful for the different language pairs. Specifically, listeners were able to discriminate between American and Australian English using '?a?a?a speech,' which contains both rhythm and intonation, but not using 'flat ?a?a?a speech,' which contains only rhythm. Thus, intonation was necessary to discriminate between the two dialects; rhythm alone was not sufficient. In contrast, listeners were unable to discriminate between American English and German using '?a?a?a speech,' but were able to discriminate the languages once intonation had been removed in the 'flat ?a?a?a speech.' Thus, the presence of uninformative intonational information apparently prevented listeners from using the informative rhythmic cues to discriminate the two languages. Once the distracting intonational information was removed, listeners could then use the existing rhythmic differences to discriminate.

Finally, previous research has shown that the ability to discriminate between prosodically similar languages is contingent on listeners' familiarity with at least one of the languages (Barkat, Ohala and Pellegrino, 1999; Ramus & Mehler, 1999). Indeed, listeners failed to discriminate either language pair using scrambled stimuli, even though

the scrambled stimuli still differed significantly in their rhythm and intonation. This supports the idea that it is easier to use prosodic information to discriminate between native/non-native language pairs than it is to discriminate between two non-native languages.

5.1.1 Limitations of the experiments on adults and future work

5.1.1.1 Do adults treat '?a?a?a speech' like speech?

The methods of re-synthesis used in Chapter 2 are similar to the re-synthesis methods used by previous researchers (Ramus & Mehler, 1999; Ramus 2002). The primary difference was that previous researchers replaced consonants with /s/, forming 'sasasa speech,' while silence was used in the current experiments. Silence was chosen over a synthesized /s/ for two reasons. In his paper, Ramus (2002) notes that participants found the constant frication of /s/ harsh sounding, and potentially distracting. Also, in Praat, it is easier to create duration-specified periods of silence than it is to create /s/ segments. Although this difference seems superficial, it is possible that, unlike 'sasasa speech,' listeners might not have treated the re-synthesized '?a?a?a speech' as speech. Specifically, it may be the case that the intervals of silence used to replace obstruents in natural stimuli are treated as pauses, rather than consonantal periods of silence.

We think this is unlikely for several reasons. Though not specifically interviewed on this matter, many participants volunteered their thoughts upon completing the

experiment. None implied they were treating the silent intervals differently than expected. If each silent interval was treated as a pause, then each vocalic interval would have been treated as a separate phrase. This seems highly unlikely, considering vocalic intervals were, on average, 104 ms (SD = 57 ms). Because silence was used to replace obstruent intervals from natural speech, their duration was usually on the order of the duration of a typical stop closure – average silent interval duration was 102 ms (SD = 63 ms).

However, whether or not listeners treat '?a?a?a speech' in the same way as 'sasasa speech' is an empirical question. Future work will attempt to replicate American English-speaking adults' ability to discriminate between 'flat ?a?a?a speech' versions of English and German using 'flat sasasa speech' where obstruent intervals are replaced with /s/, rather than silence. Listeners were able to use rhythmic differences in 'flat ?a?a?a speech' to discriminate these languages. Therefore, it is expected that they will be able to discriminate 'sasasa speech,' as well.

5.1.1.2 Can adults use only intonation to discriminate languages?

In Chapter 2, adults were tested on their ability to use only rhythm, using the 'flat ?a?a?a speech,' and their ability to use both rhythm and intonation, using '?a?a?a speech,' to discriminate languages. However, they were never directly tested on their ability to discriminate languages using intonational cues alone. The idea that intonation is weighted above rhythmic cues is based on the results from both tested language pairs –

intonation was necessary for discrimination in one case and prevented discrimination in another. Yet, in the case of American and Australian English, adults were never directly tested if they could use only intonation to discriminate dialects, or if the combination of intonation and rhythm was necessary.[10]

Testing the role of intonation alone is not straightforward, though several researchers have made attempts to do so. Ménard et al. (1999) claimed that Quebec French-listeners were relying on intonation to discriminate between low-pass filtered sentences of Quebec and European French, but this claim was based on acoustic analyses, rather than direct perceptual evidence. Willems (1982) had a trained speaker produce English sentences with English intonation or Dutch. Theoretically, these sentences were matched for segments and rhythm, but varied only in intonation (although, comparisons were not made to confirm this). The sentences used were simple, such that it was easy to produce near equivalents in Dutch. Clearly, this technique might not be feasible for other language pairs that differ in word order. It is also not clear how accurately the speaker was able to produce Dutch intonation while speaking English segments. There is also a good chance that a bilingual speaker might have a subtle accent in their pronunciation of English segments that listeners might pick up on.

Ramus and Mehler (1999) attempted to test adults' ability to use only intonation to discriminate languages by using re-synthesized stimuli - what they termed, 'aaaa speech.' To create these stimuli, they recorded sentences, and then replaced all segments

[10] This was not tested for American English and German, either. However, because listeners were only able to discriminate this language pair when intonation had been removed, it is assumed that they would fail to discriminate when given only intonation cues.

– both consonants and vowels – with /a/, and copied over the original intonational contours. Many consonants, of course, are silent, and therefore carry no pitch. To work around this, Ramus and Mehler interpolated the pitch over voiceless intervals. Thus, this method changes the original intonational cues by adding f0 contours where none previously existed. This additional information could affect listeners' ability to use intonation to discriminate languages.

I tried two alternative methods for testing the use of intonational cues directly, but they ultimately proved ineffective. For one method, I attempted to synthesize a rhythm-neutral version of '?a?a?a speech,' by randomizing the lengths of each /?/ and /a/ segment. The total number of segments remained the same, and the original pitch contours were copied over to the resulting files. Contours were either stretched or compressed to fit the duration of the new /a/ segment. However, based on acoustic analyses, these new stimuli still significantly differed in rhythm. Thus, if adults could discriminate between these re-synthesized stimuli, they may still have partially relied on rhythmic differences.

In another method, 'crossed intonation,' the intonational contour from a sentence from one language was copied onto a sentence from another. This method was only feasible for American and Australian English because the same sentences were recorded in both dialects. Thus, for each American sentence, there was a phonologically equivalent sentence in Australian, with the same number of syllables and word boundaries, allowing for a complete intonational transplant between them. The re-synthesized stimuli had original segmental and rhythmic properties, but changed intonation. Listeners were then presented with both American sentences with Australian

intonation and Australian sentences with American intonation, but listeners did not discriminate as expected. Instead, participants listened primarily to and identified sentences based on their segmental properties – American sentences were identified as American, despite their non-native intonation, nearly 100% of the time, and the same was true for Australian sentences. Adults must have been relying on segmental information in these cases because identification rates were much higher than the near-chance levels observed in the conditions when they only had access to prosodic cues.

De Pijper (1983) used a similar method to test adults' ability to discriminate languages using only intonation. He re-synthesized the intonation of recorded English sentences to have either a English-like contour or a Dutch-like contour, based on theoretical models of English and Dutch intonation. Listeners in his experiment heard the same sentences – all English, not Dutch – with either English or Dutch intonation. Thus, unlike my 'crossed intonation' experiment, listeners could not resort to segmental or rhythmic differences to identify languages, as there were only tokens from one language according to those acoustic dimensions. In that experiment, listeners could discriminate the intonation of the two languages.

Future work should test whether American English-speaking adults can use only intonation to discriminate between American and Australian English by taking sentences from one of the two dialects (ie, American English), and re-synthesizing tokens with either American or Australian intonation. Segmental information could be removed entirely by using '?a?a?a speech,' thus creating 'crossed intonation ?a?a?a speech.' These stimuli would have no segmental information, and no useful rhythmic information, as all

the rhythmic patterns would be American English. If adults could still discriminate, it would indicate intonation alone was sufficient, but if they failed to discriminate, it would imply that adults required both rhythm and intonation to discriminate American and Australian English.

5.2 Infants' discrimination ability

While infants' ability to use prosodic cues to discriminate languages has been tested for many language pairs, only one prior study has shown infants can use prosodic cues to discriminate prosodically similar languages (Bosch & Sebastián-Gallés, 1997). Researchers have previously suggested that infants primarily use rhythmic differences to discriminate between languages (Nazzi et al., 1998; 2000), but the use of different kinds of prosodic information has barely been tested (Ramus, 2002). Experiments in Chapters 3 and 4 attempted to confirm that infants can use only prosodic cues to discriminate between prosodically similar languages. They also tested which prosodic cues infants use in this task.

In Chapter 3, American English-learning infants' ability to discriminate between American English and German was tested using methods used in Nazzi et al. (2000). Using the Headturn Preference Paradigm (HPP), infants were familiarized to passages of one of the two languages, and then tested on new passages from both. Passages consisted of a subset of the sentences used with adult listeners. It was found that American English-learning 7-month-olds, but not 5-month-olds, could discriminate between

American English and German. Previous research has suggested, based on indirect evidence, that infants' language discrimination abilities improve with age. Specifically, it has been claimed that infants are born with the ability to discriminate between prosodically dissimilar languages, but are not able to discriminate between prosodically similar languages – in particular, their native language from a prosodically similar non-native language – until 4- to 5-months. Experiments in this dissertation provide the first direct evidence for this development in infants' language discrimination abilities.

Infants were also tested on their ability to distinguish between modified stimuli in order to determine which acoustic cues were used in discrimination. Infants were still able to discriminate between American English and German after the stimuli had been low-pass filtered, indicating that segmental information was not necessary for discrimination; prosodic cues alone were sufficient. Stimuli were also re-synthesized to eliminate intonational information by replacing the natural pitch contours with an artificial, sinusoidal contour. Infants were unable to discriminate between these stimuli, indicating that intonation was necessary for discrimination. Neither segmental nor rhythmic information was sufficient to support discrimination of English and German.

In Chapter 4, American English-learning infants were tested on their ability to discriminate between a number of additional language pairs, including American and Australian English, American English and Dutch, British English and Dutch, and American English and Japanese. Infants of various ages failed to discriminate between American and Australian English, and between American English and Dutch. Infants did discriminate, but showed unexpected behavior for British English versus Dutch, and

American English versus Japanese. In these latter two experiments, American English-learning 5-month-olds showed a preference for English over the non-English language regardless of the language they were familiarized to, rather than an expected novelty or familiarity effect. It was theorized that these null and unexpected preference results were due to the testing methodology. The HPP familiarization-preference design used in Nazzi et al. (2000) and this dissertation does not control for infants' habituation, so some infants may be over habituated, while others may not yet have habituated by the time they enter the test phase. To test this, American English-learning 5-month-olds were tested on British English and Dutch using an alternate paradigm, Visual Habituation (VisHab), and an alternate design which controls for habituation. In this case, infants discriminated between the two languages and showed the expected novelty effect. It was concluded that the familiarization-preference design used in the HPP experiments, though viable, is less precise and less consistent in language discrimination studies than other designs, like the habituation-novelty design used in the VisHab study.

5.2.1 Limitations of the experiments on infants and future work

5.2.1.1 Confirming that infants require intonation to discriminate

The conclusion reached in Chapter 3 – that infants rely on intonational information to discriminate their native language from a prosodically similar non-native language – is based on the fact that infants failed to discriminate between American English and

German once intonation had been removed from the stimuli and replaced with a synthesized sinusoidal contour. Infants had previously been able to discriminate between the languages using the same stimuli, either unmodified or low-pass filtered. However, there was an observed asymmetry in the fake-intonation experiment. Infants familiarized to re-synthesized German showed a significant preference to continue listening to re-synthesized German in the test phase. Infants familiarized to re-synthesized English show no language preference in the test phase.

It is unclear what exactly is causing this asymmetry, but it is clearly related to the re-synthesized intonational contour, as that is the only change between the fake-intonation experiment and the full-cue experiment. Future work could attempt to remove this confound and confirm that infants require intonation to discriminate between English and German by testing American English-learning 7-month-olds on their ability to discriminate between passages of English and German that have been re-synthesized with flat, monotone pitch.

However, as for adults, it would be desirable to test the necessity for intonational cues more directly. This is more difficult to do for American English and German, the pair needed for infants, than for American and Australian English, the pair that would be tested for adults. Unlike in the adult case, one cannot record the same sentence in both English and German without significantly changing the segmental content of the utterance. Thus, attempting to transplant the German intonation onto an English sentence could encounter problems because the same sentence might have a different number of syllables in either language. If the number of syllables between sentences were identical,

one could copy an intonational contour without losing information. However, in this case, it is possible a pitch contour indicating stress could fall on a segmentally unstressed syllable (e.g., a reduced vowel). That would certainly produce a non-native intonation in the ears of an English listener, but it would probably not be fair to still consider that contour representative of German. Therefore, one could test infants' ability to discriminate English from non-English, but not necessarily their ability to discriminate English from German. Future research would have to develop a solution to this problem before testing infants directly on their ability to use intonation to discriminate languages.

5.2.1.2 Replicating infants' discrimination patterns on a second language pair

In chapter 4, infants were tested on their ability to discriminate between American and Australian English, parallel to the adult experiments in Chapter 2. No discrimination was observed for any infant group between 5- and 9-months. However, only a small group of 9-month-olds have currently been tested. While not significant, there is a tendency for the 9-month-old group to show a novelty preference, which would indicate discrimination. In the future, I plan to test additional infants at this age in order to fill out this experimental group.

If discrimination is observed for 9-month-olds on American and Australian English, the experiments in chapter 3 can be replicated for this dialect pair. In addition, infants' ability to use only intonation to discriminate can be tested more directly with this language pair than with English and German (assuming infants can use only prosodic

information for discrimination). Specifically, a technique similar to the planned adult experiments can be used. Low-pass filtered sentences of American English can be re-synthesized to show either American or Australian English intonation.

5.3 General discussion

5.3.1 Comparing infants to adults

According to the results presented in Chapters 2 and 3, adults as well as infants can discriminate their native language from a prosodically similar non-native language using only prosodic cues. Intonation plays an important role for both ages. For adults, intonation was necessary to discriminate between American and Australian English. Similarly, for infants, intonation was needed to discriminate between American English and German.

While intonation helped, and in fact, was necessary for infants to discriminate American English and German, for adults, the presence of intonation prevented discrimination between these two languages. Once intonation had been removed, adults could use rhythmic differences to discriminate between the languages. Infants, on the other hand, could not discriminate the languages using rhythm (or rather, the combination of rhythm and segmental information). This discrepancy might be explained by the fact that for adults, when segmental information was removed, identification accuracy, as determined by A'-scores, dropped from near perfect to around 0.59. Adults rely heavily

on segmental information when identifying their native language. With infant testing procedures, there is no principled way to determine the difficulty of discrimination under different conditions – these procedures only provide information about whether or not infants are able to discriminate between languages. So we cannot comment on the relative difficulty of discrimination in the full cue versus low-pass filtered condition. Still, it is likely that as infants progress into adulthood, the reliance on segmental information increases considerably. It is possible that as attention is shifted to other domains (i.e., segmental over prosodic), adults lose the ability to discriminate between prosodically similar languages using pitch cues alone.

Another possibility is that pitch reorganizes in a way similar to segments. As infants age, they learn the phonological categories of their native language and it becomes more difficult for them to discriminate between different segments that fall within the same native phonological category. For example, English-learning 6- to 8-month-olds can discriminate between Hindi dental and retroflex stops, but 10- to 12-month-olds cannot (Werker and Tees, 1984). English-speaking adults treat both these sounds as alveolar stops. Similarly, adults may view the differences that exist between American English and German as falling within the same intonational category, making them difficult to distinguish. Seven-month-olds, on the other hand, may not have reorganized their intonational perceptual space, and therefore, would be better at discriminating these differences.

5.3.2 Reflections on linguistic rhythm

As laid out in the introduction, speech rhythm is a concept that linguists have used and discussed since Pike (1945) and Abercrombie (1967). Yet, speech rhythm has never really been adequately defined. Currently, many researchers seem to define speech rhythm as the acoustic properties captured by various rhythm metrics (Ramus et al., 1999; Grabe and Low, 2002; Wagner and Dellwo, 2004; White and Mattys, 2007). These metrics measure the patterns of segmental intervals. Ideally, these metrics are capable of classifying languages into one of three distinct rhythm-based groups: stress-timed, syllable-timed or mora-timed.

Which group a language belongs to is thought to drive how speakers of that language process speech (Mehler et al., 1981; Cutler et al., 1986; Cutler and Norris, 1988; Morais et al., 1989; Sebastian-Galles et al., 1992; Bradley et al., 1993; Pallier et al., 1993; Otake et al., 1993; Cutler & Otake, 1994; Murty, Otake & Cutler, 2007). Thus, listeners should be aware of these groups and capable of classifying languages into them, even as early as birth. This is the motivation behind the *rhythm hypothesis* (Nazzi et al., 1998; 2000).

However, researchers have pointed out several problems with the various rhythm metrics. They fail to classify non-proto-typical languages, and different metrics classify languages differently (Grabe and Low, 2002; Ramus, 2002a; Loukina et al., 2010). They are not robust in the face of varying speech rates (Ramus, 2002a; Dellwo, 2006; White

and Mattys, 2007; Arvaniti, 2009), uncontrolled speech materials (Arvaniti, 2009; Wiget et al., 2010), or inter-speaker variability (Arvaniti, 2009; Loukina et al., 2010).

In addition, little research has been done to test whether rhythm metrics measure differences that listeners actually attend to. All the debate over rhythm metrics is moot if listeners cannot use differences in segmental duration and timing to discriminate between languages. However, research in this dissertation and some earlier research (Ramus and Mehler, 1998; Ramus, 2002) show that listeners can indeed attend to differences in segment durational patterns, and use those differences to discriminate between languages. Thus, the properties of language captured by rhythm metrics can be considered important to speech perception and speech rhythm. However, speech rhythm may consist of more than segment duration and timing.

Some researchers have called for other properties of speech to be included in the definition of speech rhythm. For example, Kohler (2009) argues that the perception of 'prominence' must be important to speech rhythm. Like speech rhythm, the concept of prominence is not always well defined in linguistic research, but prominence is often connected to pitch. Stressed or pitch accented syllables – syllables marked by a particular f0 contour – are often considered to be prominent. Indeed, research in this dissertation showed that intonation plays an important role in discriminating prosodically similar languages – more important, in fact, than durational patterning. Thus, Kohler is likely right in suggesting that pitch and intonation play an important role in the perception of speech rhythm. Speech rhythm is likely constructed from both pitch and durational patterns.

However, the inclusion of pitch complicates the concept of distinct rhythm classes. For example, French and Spanish are both often classified as syllable-timed languages, yet they have very different intonational systems. Indeed, there is some evidence suggesting there may not be distinct rhythm classes at all. At the introduction of speech rhythm, languages were described as having differing amounts of both rhythmic properties: stress-timed and syllable-timed (Pike, 1945). They were not either-or classifications.

Contrary to the *rhythm hypothesis*, there are cases where infants fail to discriminate between languages that are thought to be from different rhythm classes. For instance, French-learning 2-month-olds could not discriminate between English and French, a stress-timed and syllable-timed language, respectively (Dehaene-Lambertz and Houston, 1998). English-learning 2-month-olds have been reported to fail to discriminate between French and Russian, a supposed syllable-timed and stress-timed language, respectively (Mehler et al., 1988). They also fail to discriminate between French and Japanese, the latter of which has been classified as mora-timed, and even between Dutch and Japanese, a syllable-timed and mora timed language (Christophe & Morton, 1998). Also, Mehler et al. (1988) found no evidence of discrimination between French and Russian by newborns without a French-learning background. Similarly, none of the multi-dimensional scaling studies that have developed maps of adults' similarity ratings of different languages have produced groupings that obviously align with rhythm classes (Stockmal et al., 1996; Nazzi, 1997; Bond et al., 1998; Stockmal et al., 2000; Barkat and Vasilescu, 2001; Bradlow et al., 2010).

Thus, it is likely that listeners do not classify languages into simple classes, like stress-timed or syllable-timed. Rather, listeners may classify languages in a much more complicated way, depending on the composite pitch and timing patterns in the language. In addition, these classifications are likely to be altered by the linguistic experience of the listener. Research has shown, for example, that Arabic listeners find different dialects of Arabic to be different, but non-Arabic speakers find them to be the same (Barkat, Ohala and Pellegrino, 1999). Thus, a French listener may classify some languages as distinct while an English listener classifies them as similar. In that sense, the perception of 'speech rhythm' may be very different for listeners of one language versus another.

If this is the case, then infants' language discrimination abilities become much more interesting. While adults' discrimination might be heavily influenced by their exposure to their native language, newborns and non-human animals presumably discriminate using more universal strategies. What are these universal strategies, and what similarity or language classification mapping do they produce? How does this mapping change with exposure to a single language?

5.4 Conclusion

The work presented in this dissertation demonstrated that both adults and infants can discriminate prosodically similar languages using only prosodic cues, indicating they are sensitive to very fine differences in suprasegmental cues. Specifically, intonation plays an important role in this task. Intonation was necessary for infants to discriminate, and

while adults could use only rhythmic differences to discriminate some language pairs, they weight intonation over rhythmic cues.

Bibliography

Abercrombie, D. (1967). Elements of general phonetics (Edinburgh University Press, Edinburgh).

Ananthakrishnan, S., and Narayanan, S. (2008). "Fine-grained pitch accent and boundary tone labeling with parametric F0 features " in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (Las Vegas, Nevada), pp. 4545-4548.

Arterberry, M.E. & Bornstein, M. H. (2002). Variability and its sources in infant categorization. Infant Behavior and Development, 25, 515-528.

Arvaniti, A. (2009). "Rhythm, timing and the timing of rhythm," Phonetica, 46-63.

Atkinson, K. (1968). "Language identification from nonsegmental cues," Working Papers in Phonetics, UCLA, 85-89.

Bagou, O., Fougeron, C., and Frauenfelder, U. (2002). "Contribution of prosody to the segmentation and storage of "words" in the acquisition of a new mini-language," in Speech Prosody, edited by B. Bel, and I. Marlien (Aix-en-Provence), pp. 59-62.

Bahrick, L. E., and Pickens, J. N. (1988). "Classification of bimodal English and Spanish language passages by infants," Infant Behavior and Development 11, 277-296.

Barkat, M., and Vasilescu, I. (2001). "From perceptual designs to linguistic typology and automatic language identification: Overview and perspectives.," in Eurospeech, pp. 1065–1068.

Barkat, M., Ohala, J. J., and Pellegrino, F. (1999). "Prosody as a distinctive feature for the discrimination of Arabic dialects," in Proceedings of Eurospeech '99, pp. 395-398.

Beckman, M. E., and Pierrehumbert, J. B. (1986). "Intonational Structure in Japanese and English," Phonology Yearbook 3, 255-309.

Best, C. T., Tyler, M., Gooding, T. N., Orlando, C. B., and Quann, C. A. (2009). "Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words.," Psychological Science 20, 539-542.

Bond, Z., Fucci, D., Stockmal, V., and McColl, D. (1998). "Multi-dimensional scaling of listener responses to complex auditory stimuli," in Proceedings of International Conference on Spoken Language Processing, pp. 93-95.

Bosch, L. (1998). "Bilingual exposure and some consequences on native language recognition processes at four months.," in Proceedings of the Joint Meeting of the Acoustical Society of America and the International Conference on Acoustics, edited by P. K. Kuhl, and L. Crum (Acoustical Society of America, Seattle), pp. 1599-1600.

Bosch, L., and Sebastián-Gallés, N. (1997). "Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments," Cognition 65, 33-69.

Bosch, L., and Sebastián-Gallés, N. (2001). "Evidence of early language discrimination abilities in infants from bilingual environments," Infancy 2, 29-49.

Bradley, D. C., Sánchez-Casas, R. M., and Garciá-Albea, J. E. (1993). "The status of the syllable in the perception of Spanish and English," Language and Cognitive Processes, 197-223.

Bradlow, A., Clopper, C., Smiljanic, R., and Walter, M. A. (2010). "A perceptual phonetic similarity space for languages: Evidence from five native language listener groups," Speech Communication 52, 930-942.

Bull, D., Eilers, R. E., and Oller, D. K. (1985). "Infants' discrimination of final syllable fundamental frequency in multisyllabic stimuli," Journal of Acoustical Society of America, 289-295.

Butler, J., Floccia, C., Goslin, J., and Panneton, R. (2011). "Infants' discrimination of familiar and unfamiliar accents in speech," Infancy 16, 392-417.

Byers-Heinlein, K., Burns, T. C. and Werker, J. F. (2010). "The roots of bilingualism in newborns," Psychological Science 21(3), pp. 343-348.

Choi, Y., and Mazuka, R. (to appear). "Acquisition of prosody in Korean," in Handbook of East Asian Psycholinguistics, Part III: Korean Psycholinguistics, edited by C. Lee, Y. Kim, and G. Simpson (Cambridge University Press, London).

Christophe, A., and Morton, J. (1998). "Is Dutch native English? Linguistic analysis by 2-month-olds," Developmental Science 1, 215-219.

Clarkson, M. G., and Clifton, R. K. (1985). "Infant pitch perception: Evidence for responding to pitch categories and the missing fundamental," Journal of Acoustical Society of America, 1521-1528.

Cohen, L. B., Atkinson, D. J., and Chaput, H. H. (2000). "Habit 2000: A new program for testing infant perception and cognition [Computer software]," (The University of Texas, Austin).

Collier, R., and 't Hart, J. (1981). Cursus Nederlandse Intonatie (Acco/De Horstink, Louvain).

Cooper, R. P., Abraham, J., Berman, S., and Staska, M. (1997). "The development of infant preference for motherese," Infant Behavior and Development 20, 477-488.

Cooper, R. P., and Aslin, R. N. (1990). "Preference for infant-directed speech in the first month after birth," Child Development 61, 1584-1595.

Cutler, A., and Norris, D. G. (1988). "The role of strong syllables in segmentation for lexical access," Journal of Experimental Psychology: Human Perception and Performance 14, 113-121.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1986). "The syllable's differing role in the segmentation of French and English," Journal of Memory and Language 25, 385-400.

Dainora, A. (2001). "An Empirically Based Probabilistic Model of Intonation in American English," (University of Chicago).

Dauer, R. M. (1983). "Stress-timing and syllable-timing reanalyzed," Journal of Phonetics 11, 51-62.

de Pijper, J. R. (1983). Modeling British English intonation (Foris, Dordrecht).

Dehaene-Lambertz, G., and Houston, D. (1997). "Faster orientation latencies toward native language in two-month-old infants," Language and Speech 41, 21-43.

Dellwo, V. (2006). "Rhythm and Speech Rate: A Variation Coefficient for deltaC," in Language and language processing: Proceedings of the 38th Linguistics Colloquium, Piliscsaba 2003, edited by P. Karnowski, and I. Szigeti (Peter Lang Publishing Group, Frankfurt am Main, Germany), pp. 231-241.

Diehl, R. L., Varga, K., Panneton, R., Burnham, D., and Kitamura, C. (2006). "Six-month-old infants' perception of native speech accent.," in XVth Biennial International Conference on Infant Studies (Kyoto, Japan).

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., and Vigorito, J. (1971). "Speech perception in infants," Science 171, 303-306.

Fernald, A., and Kuhl, P. K. (1987). "Acoustic determinants of infant preference for motherese speech," Infant Behavior and Development 10, 279-293.

Grabe, E., and Low, E. L. (2002). "Durational variability in speech and the rhythm class hypothesis," in Laboratory phonology, edited by C. Gussenhoven, and N. Warner (Mouton de Gruyter, Berlin), pp. 515-546.

Grice, M., Baumann, S., and Benzmuller, R. (2005). "German intonation in Autosegmental-Metrical phonology.," in Prosodic typology: The phonology of intonation and phrasing., edited by S. Jun (Oxford University Press, Oxford), pp. 55-83.

Guasti, M. T. (2002). Language acquisition: The growth of grammar (MIT Press).

Hayashi, A., Tamekawa, Y., and Kiritani, S. (2001). "Developmental change in auditory preferences for speech stimuli in Japanese infants," Journal of Speech, Language and Hearing Research 44, 1189-1200.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., and Kennedy, L. (1987). "Clauses are perceptual units for young infants," Cognition 26, 269-286.

Hohne, E. A., and Jusczyk, P. W. (1994). "Two-month-old infants' sensitivity to allophonic differences," Perception & Psychophysics 56, 613-623.

Houston-Price, C., and Nakai, S. (2004). "Distinguishing novelty and familiarity effects in infant preference procedures," Infant and Child Development 13, 341-348.

Hunter, M. A., and Ames, E. W. (1988). "A multifactor model of infant preferences for novel and familiar stimuli," in Advances in Infancy Research, edited by C. Rovee-Collier, and L. P. Lipsitt (Ablex, Stamford), pp. 69-95.

Johnson, M. H., Posner, M. I., and Rothbart, M. K. (1991). "Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging.," Journal of Cognitive Neuroscience, 335-344.

Jun, S., and Fougeron, C. (2000). "A phonological model of French intonation.," in Intonation: Analysis, Modeling and Technology, edited by A. Botinis (Kluwer Academic, Dordrecht, Netherlands), pp. 209-242.

Jun, S.-A. (2005). "Prosodic Typology," in Prosodic Typology: The phonology of intonation and phrasing., edited by S.-A. Jun (Oxford University Press), pp. 430-458.

Jusczyk, P. W., Houston, D. W., and Newsome, M. (1999). "The beginnings of word segmentation in English-learning infants," Cognitive Psychology 39, 159-207.

Karzon, R., and Nicholas, J. G. (1989). "Syllabic pitch perception in 2- to 3-month-old infants," Perception and Psychophysics 45, 10-14.

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

Kim, S. (2004). "The role of prosodic phrasing in Korean word segmentation," (UCLA).

Kim, S. and Cho, T. (2009). "The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean," Journal of Acoustical Society of America 125(5), pp. 3373-3386.

Kinzler, K. D., Dupoux, E., and Spelke, E. S. (2007). "The native language of social cognition," Proceedings of the National Academy of Sciences 104, 12577–12580.

Kitamura, C., Panneton, R., Diehl, R. L., and Notley, A. (2006a). "Attuning to the native dialect: When more means less.," in Proceedings of the Eleventh Australian International Conference on Speech Science & Technology (Australasian speech, Auckland, New Zealand), pp. 124-129.

Kitamura, C., Panneton, R., Notley, A., and Best, C. (2006b). "Aussie, Aussie Aussie, Oi-Oi-Oi: Infants love an Australian accent.," Proceedings of the conference of the Journal of the Acoustical Society of America 120, 3135.

Kohler, K. J. (2009). "Rhythm in speech and language: a new research paradigm.," Phonetica 66, 29-45.

Komatsu, M., Arai, T., and Suguwara, T. (2004). "Perceptual discrimination of prosodic types and their preliminary acoustic analysis," in Proceedings of Interspeech 2004, pp. 3045–3048.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). "Linguisitic experience alters phonetic perception in infants by 6 months of age," Science 255, 606-608.

Ladd, D. R. (1996). Intonational phonology (Cambridge University Press, Cambridge).

Lehiste, I. (1970). Suprasegmentals (MIT Press, Cambridge).

Loukina, A., Kochanski, G., Rosner, B., Keane, E., and Shih, C. (2011). "Rhythm measures and dimensions of durational variation in speech.," Journal of Acoustical Society of America 129, 3258.

Maye, J., Weiss, D. J., and Aslin, R. N. (2008). "Statistical phonetic learning in infants: facilitation and feature generalization," Developmental Science 11, 122-134.

Maye, J., Werker, J. F., and Gerken, L. (2002). "Infant sensitivity to distributional information can affect phonetic discrimination," Cognition 82, B101-B111.

Mehler, J., and Christophe, A. (1995). "Maturation and learning of language in the first year of life.," in The Cognitive Neurosciences, edited by M. S. Gazzaniga (Bradford Books, Cambridge, MA), pp. 943-954.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., and Segui, J. (1981). "The syllable's role in speech segmentation," Journal of Verbal Learning & Verbal Behavior 20, 298-305.

Mehler, J., Dupoux, E., Nazzi, T., and Dehaene-Lambertz, G. (1996). "Coping with linguistic diversity: The infant's viewpoint," in Signal to syntax: Bootstrapping from speech to grammar in early acquisition, edited by J. L. Morgan, and K. Demuth (Erlbaum, Mahwah, NJ), pp. 101-116.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halstead, N., Bertoncini, J., and Amiel-Tison, C. (1988). "A precursor of language acquisition in young infants," Cognition 29, 143-178.

Ménard, L., Ouellon, C., and Dolbec, J. (1999). "Prosodic markers of regional group membership: The case of the French of Quebec versus France," in International Conference of Phonetic Sciences (San Francisco), pp. 1601-1604.

Montgomery, C. R., and Clarkson, M. G. (1997). "Infants' pitch perception: Masking by low- and high-frequency noises," Journal of Acoustical Society of America, 3665–3672.

Moon, C., Cooper, R. P., and Fifer, W. P. (1993). "Two-day olds prefer their native language," Infant Behavior and Development 16, 495-500.

Morais, J., Content, A., Cary, L., Mehler, J., and Segui, J. (1989). "Syllabic segmentation and literacy," Language and Cognitive Processes, 57-67.

Morse, P. A. (1972). "The discrimination of speech and nonspeech stimuli in early infancy," Journal of Experimental Child Psychology 14, 477-492.

Muthusamy, Y., Barnard, E., and Cole, R. (1994). "Automatic language identification: A review/tutorial," IEEE Signal Processing Magazine 11, 33-41.

Nazzi, T. (1997). "Du rythme dans I'acquisition et le traitment de la parole," (Ecole des Hautes Etudes et Sciences Sociales, Paris, France).

Nazzi, T., and Ramus, F. (2003). "Perception and acquisition of linguistic rhythm by infants," Speech Communication 41, 233-243.

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). "Language discrimination by newborns: Towards an understanding of the role of rhythm," Journal of Experimental Psychology: Human Perception and Performance 24, 756-766.

Nazzi, T., Floccia, C., and Bertoncini, J. (1998). "Discrimination of pitch countours by neonates," Infant Behavior and Development 21, 779-784.

Nazzi, T., Jusczyk, P. W., and Johnson, E. K. (2000). "Language discrimination by English-learning 5-month olds: Effects of rhythm and familiarity," Journal of Memory and Language 43, 1-19.

Nazzi, T., Mersad, K., Iakimova, G., Polka, L., and Sundara, M. (2008). "Differences in the development of speech segmentation abilities in two French dialects.," in Acoustics'08, the 155th meeting of the Acoustical Society of America (Paris).

Ohala, J. J., and Gilbert, J. B. (1979). "Listeners' ability to identify languages by their prosody," in Problèmes de prosodie edited by P. Léon, and M. Rossi (Didier, Ottawa), pp. 123-131.

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (1993). "Mora or syllable? Speech segmentation in Japanese," Journal of Memory and Language 32, 258-278.

Pallier, C., Sebastian-Galles, N., Felguera, T., Christophe, A., and Mehler, J. (1993). "Attentional allocation within the syllabic structure of spoken words," Journal of Memory and Language, 373-389.

Panneton, R., Kitamura, C., Mattock, R., and Burnham, D. (2006). "Slow speech enhances younger, but not older infants' perception of vocal emotion," Human Development, 7-19.

Pierrehumbert, J. B. (1980). "The Phonology and Phonetics of English Intonation," (Massachusetts Institute of Technology).

Pierrehumbert, J., and Beckman, M. E. (1988). Japanese Tone Structure (MIT Press, Cambridge).

Pike, K. (1945). The intonation of American English (University of Michigan Press, Ann Arbor).

Polka, L., and Bohn, O.-S. (1996). "A cross-language comparison of vowel perception in English-learning and German-learning infants," Journal of Acoustical Society of America 100, 577-592.

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

Ramus, F. (2002). "Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues," Annual Review of Language Acquisition 2.

Ramus, F. (2002a). "Acoustic correlates of linguistic rhythm: perspectives," in Speech Prosody (Aix-en-Provence), pp. 115-120.

Ramus, F., and Mehler, J. (1999). "Language identification with suprasegmental cues: A study based on speech resynthesis," Journal of Acoustical Society of America 105, 512-521.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., and Mehler, J. (2000). "Language discrimination by human newborns and cotton-top tamarin monkeys," Science 288, 349-351.

Ramus, F., Nespor, M., and Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal," Cognition 73, 265-292.

Reitveld, A. C. M. (1980). "French word boundaries," Language and Speech 23, 289-296.

Schmale, R., and Seidl, A. (2009). "Accommodating variability in voice and foreign accent: Flexibility of early word representations.," Developmental Science 12, 583-601.

Schmale, R., Cristià, A., Seidl, A., and Johnson, E. K. (2010). "Developmental changes in infants' ability to cope with dialect variation in word recognition.," Infancy 15, 650-662.

Schneider, B., Morrongiello, B. A., and Trehub, S. (1990). "The size of the critical band in infants, children, and adults," Journal of Experimental Psychology: Human Perception and Performance, 642-652.

Schonen, S. D., Deruelle, C., Mancini, J., and Pascalis, O. (1993). "Hemispheric differences in face processing and brain maturation," in Developmental neurocognition: Speech and face processing in the first year of life, edited by B.

Boysson-Bardies, S. D. Schonen, P. W. Jusczyk, P. McNeilage, and J. Morton (Kluwer Academic Publishers, The Netherlands), pp. 149-163.

Sebastian-Galles, N., Dupoux, E., Segui, J., and Mehler, J. (1992). "Contrasting syllabic effects in Catalan and Spanish," Journal of Memory and Language, 18-32.

Seidl, A. (2007). "Infants' use and weighting of prosodic cues in clause segmentation," Journal of Memory and Language 57, 24-48.

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

Spetner, N. B., and Olsho, L. W. (1990). "Auditory frequency resolution in human infancy," Child Development, 632-652.

Stockmal, V., and Bond, Z. (2002). "Same talker, different language: A replication," in International Conference on Spoken Language Processing, pp. 77-80.

Stockmal, V., Moates, D., and Bond, Z. (2000). "Same talker, different language," Applied Psycholinguistics, 383-393.

Stockmal, V., Muljani, D., and Bond, Z. (1996). "Perceptual features of unknown foreign languages as revealed by multi-dimensional scaling," in International Conference on Spoken Language (Philadelphia, PA), pp. 1748 - 1751.

Sundara, M., Kitamura, C., and Nazzi, T. (2008). "Behavioral and physiological measures of language preference," in Acoustics '08, the 155th meeting of the Acoustical Society of America (Paris).

Thiessen, E. D., and Saffran, J. R. (2003). " When cues collide: Statistical and stress cues in infant word segmentation," Developmental Psychology 39, 706-716.

Toro, J. M., Trobalón, J. B., and Sebastián-Gallés, N. (2003). "The use of prosodic cues in language discrimination by rats," Animal Cognition 6, 131-136.

Trehub, S. E. (1973). "Infants' sensitivity to vowel and tonal contrasts," Developmental Psychology 9, 91-96.

Trehub, S. E. (1976). "The Discrimination of Foreign Speech Contrasts by Infants and Adults," Child Development 47, 466-472.

Tyler, M. D., and Cutler, A. (2009). "Cross-language differences in cue use for speech segmentation," Journal of Acoustical Society of America 126, 367-376.

Tyler, M., Perruchet, P., and Cutler, A. (2006). "A cross-language comparison of the use of stress in word segmentation," Journal of Acoustical Society of America, 3087.

Wagner, P., and Dellwo, V. (2004). "Introducing YARD (Yet Another Rhythm Determination) And Re-Introducing Isochrony to Rhythm Research," in Speech Prosody (Nara, Japan), pp. 227-230.

Ward, N. (2011). "Infants' preference for the visual speech of their native language." Unpublished M.A. thesis, UCLA.

Warner, N., Otake, T. and Arai, T. (2010). "Intonational structure as a word-boundary cue in Tokyo Japanese," Language and Speech 53(1), pp. 107-131.

Welby, P. (2003). "The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation," (Ohio State University).

Werker, J. F., and 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.

Werker, J. F., Gilbert, J. H. V., Humpfrey, K., and Tees, R. C. (1981). "Developmental aspects of cross-language speech perception," Child Development 52, 349-353.

White, L., and Mattys, S. L. (2007). "Calibrating rhythm: First language and second language studies," Journal of Phonetics 35, 501-522.

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., and Mattys, S. L. (2010). "How stable are acoustic metrics of contrastive speech rhythm?," Journal of Acoustical Society of America 127, 1559-1569.

Willems, N. (1982). English intonation from a Dutch point of view (Dordrecht, Floris).

Yoshida, K. A., Pons, F., Maye, J., and Werker, J. F. (2010). "Distributional phonetic learning at 10 months of age," Infancy 15, 420-433.