INFANT TIMBRE PERCEPTION

INFANT TIMBRE PERCEPTION: THE DEVELOPMENT OF SENSITIVITY TO SPECTRAL SLOPE

By

CHRISTINE D. TSANG, B.SC. (HONOURS)

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2003) McMaster University
(Psychology) Hamilton, Ontario

TITLE: Infant Timbre Perception: Development of Sensitivity to Spectral Slope

AUTHOR: Christine D. Tsang, B.Sc. (Honours) (McMaster University)

SUPERVISOR: Dr. Laurel J. Trainor

NUMBER OF PAGES: xii, 100

for my parents

PREFACE

This thesis consists of studies that have been previously published or have been prepared for publication as journal articles. Chapter 2 contains research that has been reprinted from *Infant Behavior and Development, Vol. 25, No. 2*, C. D. Tsang and L. J. Trainor, Spectral Slope Discrimination in Infancy: Sensitivity to Socially Important Timbres, Pages 183-194, Copyright (2002), with permission from Elsevier. The author of the present thesis is the primary author of this published work, including experimental design and programming, stimulus generation, data collection, data analysis, and manuscript preparation. Chapter 3 contains research that has been prepared and submitted for publication to the journal Developmental Science. The author of the present thesis is the primary author of this research, including experimental design and programming, stimulus generation, data collection, data analysis, and manuscript preparation. The second author of this paper is the thesis supervisor of the primary author, and the third author provided helpful assistance in data collection and analysis, as well as manuscript preparation. Chapter 4 contains research that has been prepared and submitted for publication to the Journal of the Acoustical Society of America. The author of the present thesis is the primary author of this research, including experimental design and programming, stimulus generation, data collection, data analysis, and manuscript preparation. The second author is the thesis supervisor of the primary author.

ABSTRACT

Timbre, or sound quality, is important for the identification of objects, speech sounds, people, and the vocal expression of emotion. One important cue to timbre is the spectral envelope, representing the change in intensity across frequency. Spectral slope is the linear component of the spectral envelope. Very little is known about infants' perception of timbre. However, adults are highly sensitive to very small changes in spectral shape (e.g., Green, 1992), and spectral slope is an important cue for identifying perceptual constancies in sound (Li & Pastore, 1995). Previous studies have established that infants can discriminate large differences in spectral shape (e.g., Trehub, Endman, & Thorpe, 1990; Clarkson, 1995). This thesis extends this literature by systematically examining the role of spectral slope in timbre discrimination using steady-state complex tones. The studies presented here support the idea that both infants and adults use global spectral characteristics, such as spectral slope, as a basis for timbre discrimination. The results show that 8-month-old infants are best able to discriminate spectral slope differences in sounds with moderately negative spectral slopes in the range of speech and musical sounds. Furthermore, infants show an attentional bias for sounds with slopes in this range. This thesis also demonstrates that there is a progression in the development of sensitivity to spectral slope, with initial ability to discriminate spectral slope emerging around 5 months of age. Finally, this thesis shows that spectral slope discrimination is better in adults than in infants, but that adults are similar to infants in that they are also more sensitive to spectral slopes in the moderately negative range.

ACKNOWLEDGEMENTS

Although my name is the only one on the cover, this thesis is not mine alone. Many people have contributed to the completion of this final document.

Thanks to my parents, Agnes and Ken Tsang. They have heard the good, the not-so-good, and the really, really bad. They have lived through every moment with me, and that is, perhaps, harder than doing it yourself. Thank you for worrying, for recognizing that I was upset even when I told you everything was fine, for always making sure I had nice clothes to wear to my talks, for listening to my talks over the phone, for always giving me advice and being right, for always being there—this thesis is for you.

Also, thanks to the rest of my family—my grandmother, BobbyT, Cindy, Joyce, Edmond (and Jeffrey too!). It is good to know that I have a solid family that stands behind me in everything I do.

My gratitude and appreciation to my thesis supervisor, Dr. Laurel Trainor. Thank you Laurel for taking me into your lab, and giving me an opportunity to challenge myself and see what I could do if I tried. Without your guidance, this thesis would never have been. This thesis contains only a fragment of what I have learned in your lab. Thank you to John Platt and Daphne Maurer, who have pointed me in the right direction on more than one occasion. Thanks John, for all the hours we spent on your Kyma machine trying to make my stimuli, and for all the other hours we had to spend listening to them! Thanks Daphne, for teaching me how a developmentalist thinks, and for listening when I needed you.

To Joanne Leuzzi—my lab buddy, my eyes and ears in the lab, my lunch mate, my organizer, my friend. Thanks for taking care of all the things that I never even thought of, so I wouldn't have to. To the Infant Auditory Lab past and present, especially Adrienne, Renee, Jenn, Judy, Jessica and Lisa—thanks for all the chats—they were some of the most productive days I ever had!

Special thanks to Betty Ann Levy, who was not only my department Chair, but also was my Chair for my oral defense! It was only right that you should be there at the end. Special thanks also to Kathy Murphy and Dave Jones. It's not everyone who can calculate visual angle on the phone, figure out tricky Matlab programming, and then go to the pub with you the same night!

To "the Girls" of the Claddagh ring—Carrie, Nikie, Amy, Lynne, Nicole, Sandra and Vicki. Lifelong friendships are hard to come by, but I'm fortunate enough to have 7 of them! I could not have wished for a better group of friends. Thanks for the GTs—at the Phoenix, Slainte's, the Cactus, NASA, Ottawa, New York and everywhere else we have been! It's been a great 6 years, and even more good times to come (in Brandon, Edmonton, Toronto, and wherever else we may go)!

To the McMaster Psychology Department, thanks for being such a great bunch of people! I've been lucky enough to be a part of a really good thing, and learned enough to know how rare and wonderful our department truly is.

Finally, thank you to all the parents and their babies who participated my studies—their enthusiasm and interest make developmental research possible for us all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERAL INTRODUCTION

Timbre has been described subjectively as tone quality or tone colour. In other words, timbre refers to a perceptual difference that can still be detected when all other dimensions of sound are equal. For example, timbre allows listeners to detect differences between a piano tone and a flute tone that are identical in pitch, loudness and duration. Technically, timbre is defined as "that aspect of auditory sensation by which a listener can judge two sounds that are equal in pitch, loudness and duration to be dissimilar"[1]. Timbre is an important feature of auditory objects, allowing us to identify objects, such as a flute or a piano, and specific people by the quality of their voice. Timbre is also important for communicative and social purposes. We make social judgments about others based on the quality of their voice (Bloom, Moore-Schoenmakers, & Masataka, 1999a). The discrimination of the vocal expression of emotion depends on voice timbre (e.g., Scherer, 1996; Murray & Arnott, 1993). Finally, the discrimination of speech sounds, which is crucial to spoken language comprehension, involves the perception of timbre differences.

The Dimensions of Timbre

The basis of auditory perception lies in the acoustic features of sound. There are four basic perceptual dimensions of sound: pitch, loudness, length, and timbre. Pitch is

---

[1] ANSI, Psychoacoustic Terminology, S3.20

related to the acoustic dimension of frequency, loudness is related to the acoustic dimension of intensity, and length is related to the acoustic dimension of duration. However, timbre is multidimensional. Unlike pitch, loudness and duration, timbre does not vary along a single perceptual continuum. Multidimensional scaling studies reveal that many different spectral and temporal features can contribute to perceived timbre. Despite the fact that timbre is important to auditory processing, little is known about the relative importance of the specific acoustic cues involved in timbre perception, and even less is known about how timbre perception develops.

Sounds in the world are complex in that they are composed of many different spectral frequency components or harmonics, each with a particular intensity or amplitude that can change independently over time. Both spectral and temporal characteristics contribute to the perceived timbre. For example, a change to the spectral envelope (the curve connecting the points representing the intensities of the frequency components in a tonal complex) results in a change in timbre, but in little change in perceived pitch, because, although the energy at each frequency changes, the frequency components present remain the same. As far as temporal characteristics, the onset characteristics have the most impact on timbre. In sum, the frequencies that are present, their energies, and their temporal dynamics all contribute to the perceived timbre or the quality of the sound.

The complex mapping between acoustic features and the perceived timbre makes it difficult to study. One approach that has been utilized to characterize the acoustic dimensions that are most important to timbre perception is multidimensional scaling

(MDS). Listeners are asked to rate the relative similarity between all possible pairs of stimuli in a set. The stimuli are then represented as points in a low-dimensional (usually Euclidean) space, such that the distances between stimuli reflect their relative dissimilarities. Thus, perceptually similar stimuli are represented by points clustered closely together, while dissimilar timbres are represented by points distant from one another (Kruskal & Wish, 1978). MDS is advantageous as a first pass exploratory approach because it can be used systematically to describe the similarities between natural or artificial timbres, and does not require any underlying assumptions about the nature of the dimensions in advance.

MDS has been used to scale similarities of various musical timbres (Miller & Carterette, 1975; McAdams & Cunible, 1992; Grey, 1977; Grey & Gordon, 1978) and speech sounds (deBruijn, 1978; Pols, Van der Kamp, & Plomp, 1969; Plomp, Pols, & Van de Geer, 1966). Three main characteristics of sound have been found to contribute most to perceived timbre similarities: spectral envelope, temporal envelope, and specific unique characteristics of certain sounds. Each dimension plays a role in our perception of timbre, and for some stimuli, one dimension may carry a greater weight than others. Some sounds have unique characteristics, or specificities, which are used by listeners to identify the sound. For example, clarinet tones are generally identified on the basis of their unique spectral structure, consisting of the odd-numbered harmonics, more than any specific acoustic dimension, and harpsichord tones have a characteristic "clunk" at the end of the sound (McAdams, Winsberg et al., 1995). For sounds without unique characteristics, discrimination seems to be based on spectral and/or temporal characteristics, and these two dimensions account for most of the variance in similarity judgments.

There is evidence that spectral characteristics of sounds may be weighted more heavily by listeners than temporal characteristics in most judgments of timbre similarity. When a spectral characteristic (number of frequency components) was varied compared to when a temporal characteristic was varied (shape of the spectral envelope over time), it was found that listeners judged the spectral change to be more perceptually different than the temporal change (Miller & Carterette, 1975). Other studies examining the effects of temporal vs. spectral characteristics by manipulating the temporal onsets of frequency components and maintaining the spectral structure have found that changes to the temporal structure do not significantly affect similarity judgments (Wedin & Goude, 1972). This result is consistent with the idea that spectral characteristics may play a greater role in timbre judgments.

Furthermore, the spectral centroid (the mean frequency of all the frequency components present in a complex) has also been shown to be highly correlated with judgments of timbre similarity (Grey & Gordon, 1978), supporting the notion that spectral characteristics influence timbre perception more than temporal characteristics. The spectral centroid is thought to be correlated to the subjective judgment of "brightness", in that tones with a high spectral centroid (higher representation of high frequency components) are judged "brighter" than tones with a low spectral centroid (higher representation of low frequency components). Similarity ratings between orchestral musical instrument timbres are more dependent on the centroid frequencies than on the temporal onsets of the frequency components (Iverson & Krumhansl, 1993), again suggesting that spectral composition is of greater importance to timbre judgments than temporal characteristics.

Studies using artificially generated complex tones, which have the advantage of increased control of spectral and temporal differences, have also found that the spectral envelope plays a greater role in timbre similarity judgments. When tones of equal loudness and pitch, but differing in their phase (a temporal characteristic) and amplitude pattern (a spectral characteristic), were presented to listeners, it was found that listeners judged tones with spectral differences as more different than tones with temporal differences (Plomp & Steeneken, 1969). This result is consistent with the findings of Iverson & Krumhansl (1993) and Miller and Carterette (1975).

The spectral envelope has also been found to be an important determinant for the perception and classification of vowel sounds (Plomp, Pols, & Van de Geer, 1966; Pols, Van der Kamp, & Plomp, 1969). Vowel sounds generally differ in their spectral envelopes because they have different formant frequencies (resonances in the vocal tract causing certain frequencies to be amplified in intensity). Vowels are also more affected by spectral changes than by temporal changes because they are steady-state, in that the formant frequencies are sustained and relatively unchanging across time. A multidimensional analysis of vowel spectra found that changes to the first and second formants of different vowels resulted in the greatest perceptual dissimilarity between vowels compared to changes in dynamic temporal properties (Plomp, 1966), again supporting the notion that spectral characteristics play a large role in perceived timbre differences.

It is apparent that overall spectral envelope characteristics such as the number of frequency components and spectral shape affect perceived timbre, especially for steady-state complex sounds, such as vowels. Using profile analysis, a method in which listeners detect an increase in the relative intensity of a single frequency component in the spectrum of a complex tone, it has been shown that listeners rely on spectral shape changes more than on local intensity increments for detection and discrimination (e.g., Green 1992). The human auditory system is highly sensitive to small changes in spectral shape (Green, 1992; Green & Kidd, 1982; Green, Kidd, & Picardi, 1983; Speigel, Picardi, & Green, 1981; Versfeld & Houtsma, 1991). On a subjective level, sounds which have the same frequency components but different relative intensities of those components (such as the stimuli used in profile analysis tasks) have been described by listeners to differ in tone quality, or timbre. Sounds with relatively more high frequency energy have been described as "brighter" than sounds with relatively more low frequency energy (Bismark, 1974). Versfeld & Houtsma (1991) found that listeners reported using perceived timbre changes to make discrimination judgments in a profile analysis task. Experimentally, there is much evidence to support the notion that spectral shape plays an important role in timbre discrimination. Studies of profile analysis have shown that variation in overall intensity level of complex tones has little effect on discrimination performance, but relative changes in intensity level as small as 1-2 dB in a single frequency component in a complex of over 20 components can be reliably detected

(Green, 1992; Green & Kidd, 1982; Green, Kidd, & Picardi, 1983; Speigel, Picardi, & Green, 1981; Versfeld & Houtsma, 1991). Furthermore, when the spectrum changes randomly, but the frequency component to which the intensity increment is added remains fixed, performance decreases dramatically, compared to conditions where the spectrum remains fixed but the single frequency varies (Speigel, et al., 1981), suggesting that shape of the overall spectral envelope plays a key role in the discrimination of steady state complex tones. In addition, increasing the number of components in the spectrum of a complex tone improves performance in a profile analysis task, which demonstrates that the whole spectrum is used by listeners as a cue to discrimination rather than only components near to the incremented component (Green, et al., 1983). Taken as a whole, these studies indicate that spectral shape is an important cue to timbre discrimination.

The Spectral Slope

The linear component of the spectral envelope is the spectral slope and is generally expressed in terms of decibel change per octave of frequency increase. Tones with increased representation of energy in the high frequency components have positive spectral slopes, while tones with an increased representation of energy in the low frequency components have negative spectral slopes. Figure 1 shows a real world sound spectra with a positive and a negative spectral slope. Sounds with pitch, such as vowel sounds and musical instrument tones typically have a negative spectral slope, while noise-like sounds, such as fricatives in speech sounds, have a positive spectral slope (Figure 1). Sounds with negative spectral slopes (i.e., decreasing intensity with increasing harmonic number) are much more common in the natural world than sounds

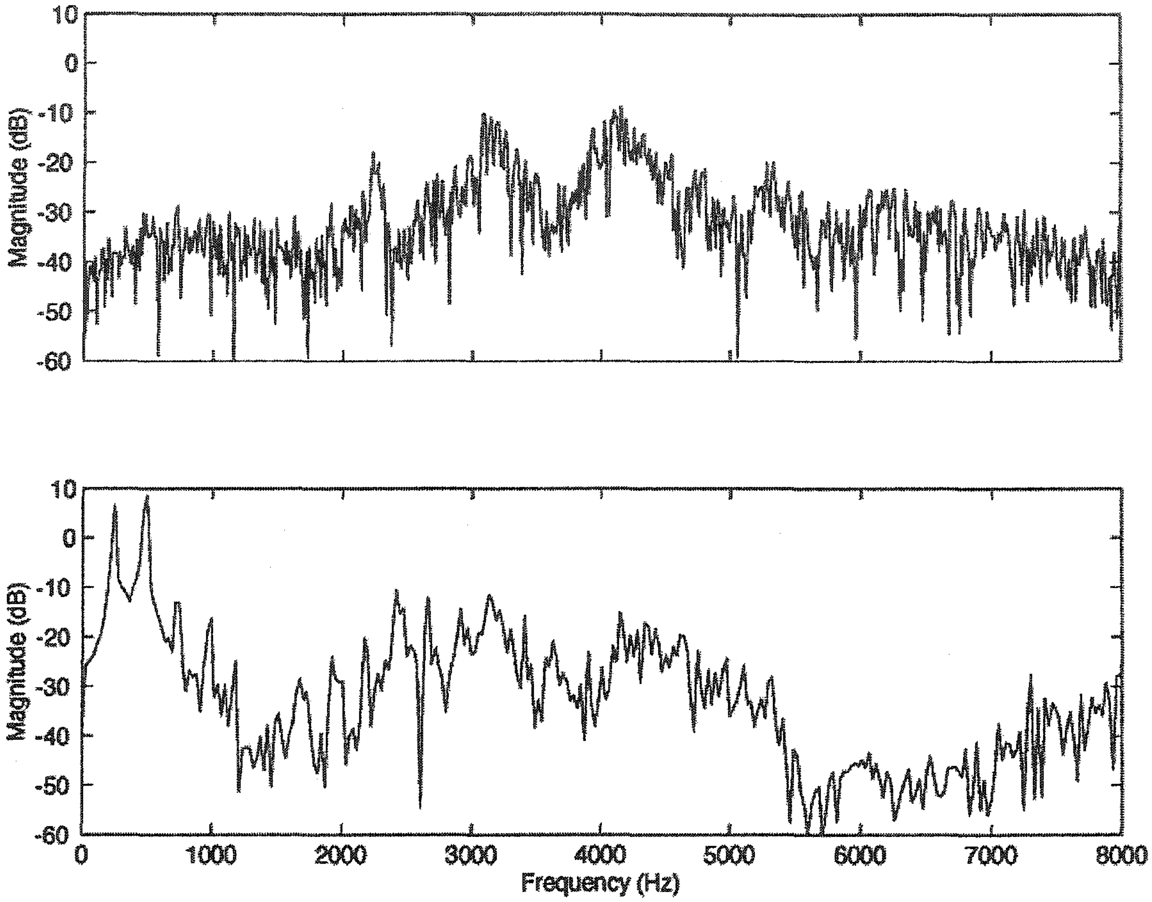


Figure 1. Sound spectra of the word "she" recorded from a female voice. The top panel shows the sound spectra of the speech fricative /sh/. The bottom panel shows the sound spectra of the vowel sound /ee/. Note the positive spectral slope at the beginning of the fricative (top panel), and the steep negative spectral slope at the beginning of the vowel sound (bottom panel).

with positive spectral slopes. The average spectral slope of the human voice has been measured to vary between about –12 and –4 dB/octave (p. 206, Hall, 1980; p.118, Sundberg, 1991). We analyzed the spectral slope of published spectra for different musical instruments (from Olson, 1967 and Fletcher & Rossing, 1991) and found that this same range also applies to musical instruments.

Spectral slope has been found to be perceptually separable from other characteristics of sound. Campbell (1994) showed that for spectral slopes between –6 and –24 dB/octave, the dimension of spectral slope was discriminated independent of temporal characteristics, such as rise time, and also to a large degree from spectral characteristics, such as cut-off frequency. Furthermore, for sounds with moderately negative spectral slopes, spectral slope discrimination does not decrease when the fine spectral structure (i.e., number and location of resonances) varies randomly, compared to conditions where the fine structure remains constant (Li & Pastore, 1995), indicating that spectral slope may be extracted independently from fine spectral characteristics. Spectral slope may be processed in parallel with other spectral characteristics, or it may be derived once fine spectral information is extracted. In any case, spectral slope appears to be perceptually independent of the fine spectral structure to a large degree.

Spectral slope is related to source characteristics of voices which remain relatively constant, while fine spectral structure is linked to filter characteristics of the vocal tract, which change from speech sound to speech sound (Li & Pastore, 1995). That spectral slope, a source characteristic, can be extracted despite variations in filter characteristics suggests that spectral slope would be an important cue for identifying perceptual constancies, such as recognizing the same voice across changing speech sounds. Spectral slope has, in fact, been implicated as a cue to a number of different auditory abilities, from voice recognition (male and female voices differ in spectral slope steepness, Klatt & Klatt, 1990), vocal quality identification (e.g., less negative (flatter) spectral slopes in the lower formants of speech sounds are correlated with highly nasal voices, Beddor, 1993; Hawkins & Stevens, 1995), and the identification of vocal expressions of emotion (e.g., expressions of negative emotions such as fear and rage are associated with flatter spectral slopes than expression of positive emotions, such as happiness, Scherer, 1989).

Adult listeners have been found to discriminate between moderately negative spectral slopes (Li and Pastore, 1995), and tilted spectra (spectra containing increased representation of either high or low frequencies) from flat spectra (spectra containing equal relative intensities across all frequency components) (Bernstein & Green, 1987). Informally, it has been reported that, with practice, discrimination of spectral slope can be as acute as 1 dB/octave for tones with negative spectral slopes (Campbell, 1994, note 3). However, we know of no study to date that has directly examined adults' thresholds for spectral slope discrimination across a wide range of spectral slope values.

Infant Timbre Perception

Before embarking on studies of infants' perception of timbre, it is necessary to establish that infants are able to hear the relevant dimensions of the sounds. While newborns' thresholds are on the order of 50 dB higher than those of adults, and they are relatively worse than adults at hearing higher frequencies, thresholds improve dramatically over the first months of postnatal life, particularly for higher frequencies, such that by 6 to 8 months of age thresholds are within 10 dB of adult levels at mid and higher frequencies and within 25 dB of adults levels for the lowest frequencies (see Werner & Marean, 1996 for a review). Thus, in the second half year of life, it is not necessary to play sounds very loudly in order for infants to hear them.

Frequency discrimination is also reasonably good in infants. At 3 months, infants are able to discriminate frequency differences of 3% to 4% under conditions where adults can discriminate changes of 1% (Olsho, Koch, & Halpin, 1987). By 5 to 9 months of age, frequency discrimination has improved markedly, with infants showing adult thresholds at 4000Hz, and responding to 2% changes in the frequency of 500 Hz tones under conditions where adults respond to changes of 1% (Olsho, Shcoon, Sakai, Turpin, & Sperduto, 1982a, b; Sinnott & Aslin, 1985). Thus, by the second half year of life, infants are quite good at frequency discrimination. Less is known about infants' ability to discriminate intensity differences. However, Sinnott and Aslin (1985) reported thresholds in 7- to 9-month-olds of about 6 dB for a 1000 Hz tone burst compared to adult thresholds of less than 2 dB. Bull, Eiler, & Oller (1984) found that infants showed some response to a 2 dB increase in the final syllable of 3-syllable stimuli and a robust response to a 4 dB increase. Thus, infants between 5 and 8 months of age are able to discriminate intensities, although more research is needed in order to determine the factors that affect their thresholds. In sum, by the second half year of life, infants certainly have the frequency and intensity discriminative ability necessary for timbre discrimination.

There have been few studies directly examining infant timbre perception. However, timbre is a distinguishing feature of infant-directed singing, in that adult listeners rate infant-directed singing as being rendered in a "more loving tone of voice" than non-infant-directed singing. Infants listen longer to the infant-directed versions, and there is a correlation between adults' ratings of how loving the voice is and infants' listening times (Trainor, 1996), a result suggesting that infants prefer the timbre of infant-directed versions in comparison to that of non-infant-directed versions. There is indirect evidence that even very young infants use timbre information. Two-day-old infants recognize their mother's voice (DeCasper & Fifer, 1980), and 2-month-old infants are able to discriminate between male and female voices (Miller, 1983). However, cues to discrimination other than spectral differences are present in these stimuli, such as pitch and talker-specific intonation patterns. However, 2- and 3-month-old infants have been found to discriminate between vowel categories despite variability in pitch and talker (Marean, Werner, & Kuhl, 1992; Kuhl, 1979), suggesting that infants are using global properties of the sound, such as the overall spectral envelope, as a basis for discrimination. Together, this evidence supports the notion that timbre cues can provide infants with information regarding perceptual constancies in sound (e.g., the same voice saying different vowels), as well as providing cues about differences between sounds (e.g., identifying different voices saying the same vowel). Thus, timbre is important to the processes of both discrimination and recognition.

Very little is known about the development of the ability to detect differences in the spectral envelope. That infants are able to categorize vowel sounds across different talkers (Kuhl, 1979) is indirect evidence that infants are using spectral envelope properties in their auditory processing. By 7- and 8-months of age, infants have been found to discriminate vowel-like sounds that differ only in the shape of the spectral envelope (Clarkson, Clifton, & Perris, 1988; Trehub, Endman, & Thorpe, 1990). More directly, Clarkson (1996) showed that 7-month-old infants can discriminate between two highly different spectral profiles, a single negatively sloped spectrum and a single positively sloped spectrum, despite a varying (or roving) overall intensity level, a result suggesting that infants are also sensitive to spectral envelope properties of sounds.

However, how sensitive infants are to changes in the spectral envelope is not known. The importance of timbre in auditory processing and the importance of global properties of the spectrum in timbre perception motivate an examination of infants' sensitivity to changes in spectral shape. This thesis focuses on the role of spectral slope in the perception of the timbre of complex tones. In Chapter 2, 8-month-old infants' discrimination of a wide range of spectral slopes was measured, first, to examine whether infants are sensitive to small differences in spectral slope and, second, to test whether infants are more sensitive to the range of spectral slopes most commonly found in speech and musical sounds. Chapter 3 explores this question in younger 5-month-old infants, and Chapter 4 explores it in adults. Chapter 3 also examines the question of whether infants attend longer to complex tones with spectral slopes in the range of speech and music in comparison to tones with spectral slopes outside this range. This question is important because infants need to attend to speech and music in order to maximize their learning of these important structures. Finally, Chapter 5 summarizes the results and examines possible causes of the developmental changes observed.

CHAPTER 2: INTRODUCTION

As discussed in Chapter 1, the spectral envelope plays a large role in the perception of the timbre of complex sounds, such as voices, speech sounds and musical instruments. While adults have been shown to be highly sensitive to changes in spectral shape (e.g., Green, 1983), very little is known about infant sensitivities. That infants can discriminate between voices and speech sounds provides indirect evidence of the ability to discriminate sounds based on global spectral differences. Clarkson, Clifton, and Perris (1988) and Trehub, Endman, and Thorpe (1990) have shown that 7- and 8-month-old infants can categorize vowel-like sounds differing only in the shape of the spectral envelope. A more direct examination of infants' discrimination of spectral shape found that 7-month-olds can discriminate between highly different spectral profiles despite varying overall intensity level (Clarkson, 1996), a finding showing that by 7 months, infants are able to use global spectral shape properties as cues for discrimination.

Most sounds in the natural world have moderately negative spectral slopes. The previous studies in the literature have established that infants are sensitive to changes in spectral shape, but it is not known whether infants have equal sensitivity across the range of spectral slopes, or if infants will show enhanced discrimination abilities in the range of real world sounds. This chapter will examine whether 8-month-old infants are sensitive to small differences in spectral slope, and also determine if sensitivity is greatest in the range of spectral slopes most commonly found in speech and music.

Spectral Slope Discrimination in Infancy: Sensitivity to Socially Important Timbres

Christine D. Tsang & Laurel J. Trainor

McMaster University

Running head: spectral slope discrimination in infancy

Address correspondence to:
Christine D. Tsang
Department of Psychology
McMaster University
Hamilton, ON
Canada, L8S 4K1
905-525-9140 ext. 27114
fax: 905-529-6225
email: tsangcd@mcmaster.ca

Acknowledgements

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Laurel J. Trainor. We wish to thank John Platt, Terri Lewis and Kathleen Bloom for their comments on an earlier version of the manuscript and Joanne Leuzzi for her invaluable assistance in the lab.

Abstract

Spectral slope, the linear component of the spectral envelope, affects sound quality or timbre, and is important for object identification, speech discrimination, voice recognition, and interpreting vocal expressions of emotion. Eight-month-old infants discriminated between tones with spectral slopes of –10 dB/octave and –4 dB/octave, but not between positive or highly negative spectral slopes. Thus, the infant auditory system is tuned to be most sensitive to spectral slope differences in the range of those commonly found in speech and music.

Key words: auditory, infants, timbre, spectral slope, discrimination

One important source of information about objects in the world is the sounds they emit. Sounds can be distinguished on the basis of pitch, duration, loudness, and sound quality or timbre. While pitch, loudness, and duration can be specified largely on the basis of a single physical dimension (i.e., frequency, sound pressure levels or intensity, and time, respectively), no single physical dimension fully characterizes timbre. In fact, timbre is defined negatively as "that aspect of auditory sensation by which a listener can judge two sounds that are equal in pitch, loudness and duration to be dissimilar" (ANSI, 1973, Psychoacoustic Terminology, s3.20). For example, if the same note is played on a flute and a piano for the same length of time at the same loudness, timbre is what differs between them. From a social perspective, timbre is vitally important. Speech sounds or phonemes are distinguished largely on the basis of timbre, and we recognize different speakers by the quality of their voice. Furthermore, timbre plays a large role in the vocal expression of emotion, and adults can recognize emotions in speech across languages and cultures (Scherer, 1989; Frick, 1985; Murray & Arnott, 1993). In this paper, we ask whether the infant auditory system is tuned to process aspects of timbre that are important in speech and music perception.

Infants certainly process some aspects of timbre. They can recognize their mother's voice at 2 days of age (DeCasper & Fifer, 1980). They can distinguish speech sounds early in life (Kuhl, 1979). Speech directed to infants tends to be emotionally expressive (Fernald, 1991; Trainor, Austin, & Desjardins, 2000) and infants as young as 5 months of age respond differently to different emotional messages (Fernald, 1993; Rock, Trainor, & Addison, 1999; Walker-Andrews & Grolnick, 1983). It should be noted, of course, that in both voice and emotion recognition, changes in pitch and prosody may also serve as a cues for discrimination. However, it is clear that at least as young as 7 months, infants can discriminate sounds differing only in timbre (Clarkson, Clifton, & Perris,

1988) and categorize speech-like sounds by timbre (Trehub, Endman, & Thorpe, 1990). Yet we understand little about the particular cues infants use for timbre discrimination.

Most sounds in the world are complex, that is, they are composed of several frequency components or harmonics. These components may differ in relative intensity or amplitude. The spectral envelope is defined as the curve that connects the points representing the amplitudes of the frequency components in a tonal complex (Moore, 1997). Studies using multidimensional scaling of adult perceptions of steady-state complex stimuli, based on pairwise timbre similarity judgements, show that perceived differences in the timbre of steady-state sounds depend largely on characteristics of the spectral envelope (Wedin & Goude, 1972; deBruijn, 1978).

By manipulating the relative intensities of the harmonics (frequency components) that make up a stimulus, the shape of the spectral envelope can be changed, while keeping the frequencies of the harmonics (and therefore to first approximation, the pitch) the same. In other words, the relative intensities at each of the frequency components in a complex tone can be changed, resulting in a change in timbre, but not in pitch. These changes to the spectral envelope are generally perceived by adult listeners to be a change in the sound quality or timbre. Adults are very good at discriminating small changes in the spectral envelope (e.g. Green, 1983). Infants have also been found to be able to detect changes in the spectral envelopes of complex steady-state tones. Seven-month-old infants discriminate between tonal complexes that have the same fundamental frequency, but differ in their harmonic components (Clarkson et al., 1988). Trehub et al. (1990) found that 7- to 8.5-month-old infants are able to differentiate between spectral structures of complex tones despite variation in fundamental frequency, intensity and duration. Furthermore, 7-month-old infants are also able to discriminate sounds with highly different spectral profiles (i.e., rising vs. falling spectrum) (Clarkson, 1996).

In the present paper, we concentrate on the *spectral slope*, which is a global property of the spectral envelope representing the linear component of change in relative intensity level across spectral frequency. Most sounds in the natural world, including voices, have negative spectral slopes (i.e. intensity decreases with increasing harmonic number), and thus low frequency components have relatively more intensity than the high frequency components. Spectral slope is important in object identification, the perception of speech sounds and the vocal expression of emotion. Female voices have been found to have flatter spectral slopes than male voices, resulting in relatively greater intensities at the high frequencies than in male voices (Hattaori, Yamamoto, Fujimura, 1958; Monsen & Engebretson, 1977; Huffman, 1990; Klatt & Klatt, 1990; Nittrouer, McGowen, Milenkovic, & Beehler, 1990). The perception and identification of vowels can also be affected by spectral slope steepness (Lea & Summerfield, 1994). Flatter spectral slopes (i.e., less negative spectral slopes, and therefore relatively more intensity at high frequencies) have been found for voices expressing fear and rage, whereas steeper negative spectral slopes (i.e., relatively less intensity at high frequencies) have been found for vocal expression of happiness (Scherer, 1989).

Spectral slope is also implicated in our social impressions of others. Sounds that have more intensity at high frequency components (i.e., less negatively sloped or positively sloped) are perceived to sound nasal. Nasality influences social attitudes and perceptions, such that increased nasality is associated with negative characteristics, such as weakness (Bloom, Zajac, & Titus, 1999) and "whining" (Laver, 1980). Female voices are generally more nasal than men's voices (Seaver, Dalston, Leeper, & Adams, 1991). However, mothers' infant-directed singing is less nasal than their non-infant-directed singing in that the former contains relatively more intensity at the low-frequencies than the latter (Trainor, Clark, Huntley, & Adams, 1997). Nasality also affects adults' social perceptions of infant vocalizations. Vocalizing 3-month-old male infants are rated as more socially favorable than female infants, even when the sex of the infant is not revealed (Bloom, Moore-Schoenmakers, & Masataka, 1999). The only acoustic difference between male and female infant vocalizations appears to be higher nasality in the female vocalizations. This has important social implications because mothers have been observed to be less responsive to infants producing sounds of greater nasality (Masataka & Bloom, 1994).

In sum, spectral slope perception is important in object identification and speech perception as well as social and emotional interaction. We ask whether young infants are sensitive to spectral slope. The average spectral slope of the human voice has been measured to be between about –12 and –4 dB/octave (p. 118, Sundberg, 1991; p.206, Hall, 1980). We analyzed the spectral slope of published spectra for different orchestral instruments (from Olson, 1967 and Fletcher & Rossing, 1991) and found that this same range applies to musical instruments as well. Previous research has established that 7-month-old infants are able to discriminate a single positive spectral slope (linear increase of 3 dB per 200 Hz) from a single negative spectral slope (linear decrease of 3 dB per 200 Hz) (Clarkson, 1996). Infants are particularly interested in speech and musical sounds. Since the majority of these meaningful sounds in the environment have spectral slopes in the moderately negative range, enhanced sensitivity for spectral slopes in this range would be expected if spectral slope is important for infants' discrimination of timbre. In the present experiment, 8-month-old infants' discrimination of a wide range of spectral slopes was measured to test whether the developing auditory system is differentially tuned to specific ranges of spectral slopes.

Method

Participants

Forty-one (30 male, 11 female) normal, full-term 8-month-old infants (*M* = 33. 9 weeks, range = 32.0 – 36.1 weeks) participated. All infants were born within 2 weeks of full term and were healthy at the time of testing, had no history of chronic ear infections or suspicion of hearing loss. Infants were randomly assigned to one of five discrimination conditions until 8 infants in each condition passed from Phase 1 to Phase 2 of the testing procedure (see Procedure) or a maximum of 11 infants had been tested in each condition. Table 1 shows the total number of infants tested in each condition.

A further 14 infants (3 male, 11 female, M = 33.5 weeks, range = 32.0 – 36.1 weeks) were tested in one condition in order to equate the number of infants reaching criterion (see Results).

Stimuli

Complex tones were generated on a Macintosh IIci computer using Synthesize software. All tones had a fundamental frequency of 200 Hz (this is in the range of female voices) and were 1000 ms in duration, including linear 20-ms rise and fall times. Each tone consisted of the first five harmonics in cosine phase. Seven different spectral slopes were used: –16, –10, –4, –3, +3, +4, and +16 dB/octave. Each slope was presented at five different intensity levels (65, 67, 69, 71 and 73 dB(A) over a noise floor of 24 dB(A)) in order to minimize infants' use of local intensity cues. Discrimination of four pairs of spectral slopes were tested (see Table 1). As can be seen in Table 1, larger spectral slope differences were tested for positive than for negative spectral slopes, as discrimination was expected to be poorer for the positive than for the negative slopes.

Table 1.
Spectral slope pairs tested, the number of infants tested on each pair, and the number reaching criterion in Phase 1.

	slope 1 (dB/octave)	slope 2 (dB/octave)	number of infants tested	number of infants reaching criterion	proportion of infants reaching criterion
i)	–16	–10	11	5	0.45
ii)	–10	–4	8	8	1.00
iii)	–3	+3	11	2	0.18
iv)	+4	+16	11	4	0.36

Apparatus

Testing was conducted in a sound-attenuating chamber (Industrial Acoustics Co.). A Macintosh IIci computer with an audiomedia card generated the 16-bit sounds and ran the experiment. The sound stimuli were passed through a Denon amplifier (PMA-480) to a single loudspeaker (GSI) located inside the booth. The parent holding the infant sat in a chair arranged so that the loudspeaker was located on the infants' left. Under the loudspeaker were four compartments, each containing lights and a mechanical toy. It was not possible to see into the compartments except during reinforcement (see Procedure). The experimenter was seated facing the infant.

Procedure

Infants were tested individually in a go/no-go conditioned head-turn response procedure (e.g., Trainor & Trehub, 1992). The experimenter and parent listened to masking music through headphones so as to be unaware of what the infant was hearing. During the experiment, one of the spectral slopes in a pair (see Table 1), the standard spectral slope, repeated continuously from a loudspeaker on the infant's left. The interval between the tones was 1000 ms. When the infant was attentive and facing the experimenter, a trial was initiated by the experimenter. There were two types of trials: control trials, in which a tone with the standard spectral slope was presented, and change trials, in which a tone with a different spectral slope was presented. If the infant made a 45 $^{\circ}$ or greater head turn (as judged by the experimenter after training) toward the loudspeaker within 3000 ms of the beginning of a change trial, a toy in one of the compartments under the loudspeaker lit up for 3000 ms as a reinforcer. Head turns at other times and those that were less than 45° were not reinforced. Once the lights were extinguished, the experimenter attracted the infant's attention forward again. The computer kept track of any head turns that occurred within a 3000 ms window on change trials (hits) as well as on control trials (false alarms) to provide an index of the rate of random turning. Trials were presented in a quasi-random order for each subject, with the constraint that no more than two control trials occurred sequentially.

Each infant was tested with one of the pairs of slopes shown in Table 1. For half of the infants, a tone with slope of *Slope 1* (see Table 1) was presented as the standard and a tone with *Slope 2* (see Table 1) was presented during change trials. For the other half of the infants, this was reversed. There were two experimental stages, *Phase 1* and *Phase 2*. During Phase 1, only change trials were presented and intensity was held constant at 71 dB(A) to make the task easier. Demonstration trials, in which the change slope was presented paired with the activation of a toy, were presented if the infant failed to turn on several trials in a row, in order to show that head-turning to a change tone would be rewarded. A criterion was set at 4 correct trials in a row within 20 trials. If the infant failed to reach criterion, the session was terminated. If the infant reached criterion, Phase 2 began. During this phase, both the standard and change tones were presented with the full range of intensity levels (65 to 73 dB(A)), with the intensity on a particular trial chosen randomly. There were no demonstration trials during Phase 2 . Each infant completed 24 trials in Phase 2: 12 change trials and 12 control trials. The same procedure was followed for all slope pairs.

Results

There were no significant differences in performance in any condition depending on whether Slope 1 or Slope 2 was the background, so the data were collapsed across this variable.

All of the 8 infants tested in the –4/–10 discrimination condition successfully reached criterion, but at most only 5 out of 11 infants tested in each of the other conditions did so (Table 1). A 2 (pass or fail) x 4 (condition) chi square analysis was conducted on the number of infants passing or failing in Phase 1 of each condition, $x^2(3)= 16.955$, $p < 0.001$. Thus, the null hypothesis that there were equal proportions of infants passing and failing in each condition was rejected. A further set of chi-square analyses were conducted using pair-wise comparisons between each of the conditions, to determine which conditions differed significantly. Comparisons between the –4/–10 condition and each of the other conditions yielded significant results (x^2 (1)=10.588, $p<0.001$ for –4/–10 and –10/–16; x^2 (1)=10.588, $p<0.001$ for –4/–10 and –3/+3; x^2 (1)=15.856, $p=0.000$ for –4/–10 and +4/+16). Comparisons between all other pairs of conditions were not significant. Thus, when there was no intensity variation, the –4/–10 condition was relatively easy compared to the other conditions.

In phase 2, infants in the –4/–10 condition responded correctly (hits) on 0.47 proportion of trials and responded incorrectly (false alarms) on 0.35 proportion of trials, which differ significantly from each other according to a paired sample t-test, $t(7)=3.52$, $p<0.005$. The results are the same when d' is used. A single-sample t-test showed that discrimination between the tones was significantly above chance levels, $d' = 0.32$,

SD=0.26, $t(7)$=3.52, $p < 0.005$, indicating that infants were able to discriminate spectral slopes in this moderately negative range[1]. Because so few subjects were able to reach criterion in the other conditions, it was not possible to conduct statistical tests. However, the mean *d'*s and percents correct (based on change and control trials) in each case were very low for the few infants tested in Phase 2 (*d'*=0.2025, proportion of hits = 0.5, proportion of false alarms = 0.42 for –16/–10; *d'*= -0.09, proportion of hits = 0.50, proportion of false alarms = 0.54 for –3/+3; *d'*= 0.15, proportion of hits = 0.52, proportion of false alarms = 0.46 for +4/+16). Thus, infants were only able to discriminate spectral slope in the –4/–10 case. Even when there was no intensity rove (Phase 1), infants found spectral slope discrimination difficult outside the –4/–10 dB/octave range.

Because few infants met the criterion in Phase 1 on the –16/–10, –3/+3, and +4/+16 conditions, few infants were tested in Phase 2 (with the intensity rove) in these conditions. While it is unlikely that infants would perform well in the more difficult Phase 2 task if they had difficulty in Phase 1, we decided to test this in one of the conditions. Thus, a further 14 infants were tested in the +4/+16 condition in an effort to match the number of infants reaching criterion to that of the –4/–10 condition. A total of 9 of the 25 infants were able to reach criterion in Phase 1 and completed Phase 2. A single-sample *t*-test on the *d'* values from Phase 2 found no significant difference from chance levels, mean *d'*=0.04, SD=0.39; $t(24)$=0.33, p=0.37. A paired *t*-test on the proportion of hits (0.52) and false alarms (0.50) in Phase 2 also found no significant

---

[1] The *d'* values are generally quite low, even in the –4/–10 condition. The variation in intensity from trial to trial likely makes this task very difficult. It must be noted that the performance of prelinguistic infants in psychophysical tasks rarely reaches adult levels. This is likely due to motivational and attentional factors, in conjunction with the lack of verbal instructions. The important result is that performance in the –4/–10 condition was well above chance levels.

differences, $t(8)=0.34$, $p>0.7$. Thus, even the top-performing 36% of infants in the +4/+16 condition did not show discrimination between the pair of tones, suggesting that this condition is much more difficult for infants than the –4/–10 condition, in which infants showed discrimination, even though 100% of infants were included.

Theoretically, discrimination between two tones of differing spectral slope could be based on listening for "local" changes that occur in a single harmonic within the complex, rather than attending to the whole complex. Specifically, a listener could "hear out" a single harmonic in the complex, and use intensity changes that occur in this single harmonic to make the discrimination between a pair of spectral slopes. A set of calculations (see Appendix A) was carried out to determine whether the intensity variation of 8 dB was sufficient to control for the use of local intensity differences. Each pair of spectral slopes contained the same 5 harmonics, but at different intensities. For example, tones with highly negative spectral slopes have more relative intensity at low frequency components than slopes with less negative or positive spectral slopes. In order to determine how large the difference was between the equivalent harmonics of any given pair of spectral slopes tested, local intensity cues between each pair of spectral slopes were calculated. In other words, the difference in intensity between each of the corresponding harmonics of the two complexes was determined. The maximum difference in intensity (the value of the largest local cue) across corresponding harmonics for each pair was considered to be the intensity variation required to mask local intensity cues (See Figure 1 in Appendix A). Although the local intensity differences (i.e. differences at individual harmonics) between spectral slope pairs exceeded the intensity variation used during the test phase (Table 2), the greatest local intensity cues were in the +4/+16 condition, where performance was very poor. This suggests that infants were not attending to local intensity cues, but were attending to the global spectral envelope.

Table 2.
Intensity rove needed to eliminate use of local intensity cues for each pair of spectral slopes.

slope 1	slope 2	intensity rove needed (dB)
–16	–10	13.48
–10	–4	11.70
–3	+3	8.15
+4	+16	25.00

The data were also examined by intensity to determine if performance in the –10/–4 condition could have been due to infants responding only on the most intense trials (i.e., 73 dB(A)). A one-way repeated measures ANOVA on the proportion of head-turns in Phase 2 within each intensity condition showed that there was no significant effect of intensity on performance, indicating that infants were not responding to differences in loudness, but rather to differences in tone quality.

Discussion

Eight-month old infants appear to be most sensitive to differences in spectral slope in a limited range around –10 and –4 dB/octave. Infants did not show discrimination between the highly negative spectral slope pair or between the positive spectral slope pair, even though the latter difference was large (+4/+16 condition). While infants were shown previously to discriminate a single highly negative spectral slope from a single highly positive spectral slope (Clarkson, 1996), the results of the present study extend Clarkson's (1996) findings by suggesting that infants are maximally sensitive to spectral slope differences in the moderately negative region that are contained in speech (p. 118, Sundberg, 1991; p.206, Hall, 1980) and music (Olson, 1967; Fletcher & Rossing, 1991).

The results are particularly important because they suggest that 8-month-old infants are able to use the global component of spectral envelope to process and organize auditory input that is relevant to the real world tasks of object recognition, vocal expression of emotion, and social interaction. Spectral slope differs between male and female voices (e.g., Klatt & Klatt, 1990), and may be the basis by which infants are able to discriminate male from female voices matched in pitch (Miller, 1983). Infants' sensitivity to spectral slope differences in the range of human voices also suggests that infants are likely to be sensitive to socially relevant differences in vocal nasality, which depend crucially on spectral slope discrimination.

Differences in spectral slope may also contribute to infants' ability to discriminate their mother's voice from that of a stranger's at 2 days of age (DeCasper & Fifer, 1980), although other cues such as pitch and prosody must also play a large role in these types of discriminations. It should be noted that timbre is a highly complex dimension of sound, and it is likely that spectral slope and other steady-state characteristics of the spectral envelope are not the only relevant acoustic components in timbre discriminations. Indeed, dynamic aspects of the spectral envelope play a crucial role in the identification and recognition of consonants in speech for adult listeners (Liberman, Harris, Kinney, & Lane, 1961; Cutting & Rosner, 1974) as well as infants (Jusczyk, Rosner, Cutting, Foard, & Smith, 1977). Temporal characteristics of sound are also an important distinguishing feature in musical instrument timbres (Grey, 1977; McAdams, Winsberg, Donnadieu, De Soete, & Krimpoff, 1995; Cutting and Rosner, 1974).

Nonetheless, it is clear from the results presented here that infants are at least sensitive to the global spectral envelope.

It remains for future research to test whether infants younger than 8 months show sensitivity to spectral slope differences relevant to speech and music, as well as whether this sensitivity generalizes to sounds with more complex spectral envelopes, such as voices. More generally, this study did not directly address the role of experience in spectral slope discrimination. Thus, it is not known whether the enhanced discrimination for real world spectral slopes is due to exposure to real-world sounds, or whether this sensitivity has been specified by genetics. If experience is the main cause, it might be predicted that increased exposure to sounds with positive spectral slopes would increase the discriminability of positive spectral slopes. On the other hand, if discriminability of spectral slopes is genetically specified, further artificial exposure to environmentally rare spectral slopes would not be expected to modify spectral slope discriminability substantially.

Whatever the cause, from an infant's perspective, it would certainly be advantageous to be able to discriminate timbres that signal important social messages. By the measure of spectral slope discrimination, infants are tuned to be sensitive to timbre differences important for object recognition, speech discrimination and the vocal expression of emotion.

References

Bloom, K., Moore-Schoenmakers, K. & Masataka, N. (1999). Nasality of infant vocalizations determines gender bias in adult favorability ratings. Journal of Nonverbal Behavior, 23, 219-235.

Bloom, K., Zajac, D.J., & Titus, J. (1999). The influence of nasality of voice on sex-stereotyped perceptions. Journal of Nonverbal Behavior, 23, 271-281.

deBruijn, A. (1978). Timbre classification of complex tones. Acustica, 40, 108-114.

Clarkson, M.G. (1996). Infants' intensity discrimination: Spectral profiles. Infant Behavior and Development, 19, 181-190.

Clarkson, M.G., Clifton, R.K., & Perris, E.E. (1988). Infant timbre perception: Discrimination of spectral envelopes. Perception and Psychophysics, 43, 15-20.

Cutting, J.E., & Rosner, B.S. (1974). Categories and boundaries in speech and music. Perception and Psychophysics, 16, 564-570.

DeCasper, A.J., & Fifer, W.P. (1980). Of human bonding: Newborns prefer their mothers' voices. Science, 208, 1174-1176.

Fernald, A. (1991). Prosody in speech to children: Prelinguistic and linguistic functions. Annals of Child Development, 8, 43-80.

Fernald, A. (1993). Approval and Disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. Child Development, 64, 657-674.

Fletcher, N.H., & Rossing, T.D. (1991). The Physics of Musical Instruments. New York: Springer-Verlag.

Frick, R.W. (1985). Communicating emotion: The role of prosodic features. Psychological Bulletin, 97, 412-429.

Green, D.M. (1983). Profile analysis: A different view of auditory intensity discrimination. American Psychologist, February, 133-142.

Grey, J.M. (1977). Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America, 61, 270-277.

Hall, D.E. (1980). Musical Acoustics: An Introduction. Belmont, California: Wadsworth Publishing Co.

Hattatori, S., Yamamoto, K., & Fujimura, O. (1958). Nasalization of vowels in relation to nasals. Journal of the Acoustical Society of America, 30, 267-274.

Huffman, M.K. (1990). Implementation of Nasal: Timing and Articulatory Landmarks. A UCLA Working Paper in Phonetics.

Jusczyk, P.W., Rosner, B.S., Cutting, J.E., Foard, C.F., & Smith, L.B. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. Perception & Psychophysics, 21, 50-54.

Klatt, D.H. & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America, 87, 820-857.

Kuhl, P.K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America, 66, 1668-1679.

Laver, J. (1980). The Phonetic Description of Voice Quality. Cambridge: Cambridge University Press.

Lea, A.P. & Summerfield, Q. (1994). Minimal spectral contrast of formant peaks for vowel recognition as a function of spectral slope. Perception and Psychophysics, 56, 379-391.

Liberman, A.M., Harris, K.S., Kinney, J.A., & Lane, H. (1961). The discrimination of relative onset time of the components of certain speech and nonspeech patterns. Journal of Experimental Psychology, 61, 379-388.

Masataka, N. & Bloom, K. (1994). Acoustic properties that determine adults' preferences for 3-month-old infant vocalizations. Infant Behavior and Development, 17, 461-464.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimpoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research, 58, 177-192.

Miller, C.L. (1983). Developmental changes in male/female voice classification by infants. Infant Behavior and Development, 6, 313-330.

Monsen, R.B. & Engebretson, A.M. (1977). Study of variations in the male and female glottal wave. Journal of the Acoustical Society of America, 62, 981-993.

Moore, B.C. (1997). An Introduction to the Psychology of Hearing (4th ed). London: Academic Press.

Murray, I. R. & Arnott, J.L. (1993). Toward simulation of emotion in synthetic speech, A review of the literature on human vocal emotion. Journal of the Acoustical Society of America, 93, 1097-1108.

Nittrouer, S., McGowan, R.S., Milenkovic, P.H., Beehler, D. (1990). Acoustic measurements of men's and women's voices: A study of context effects and convariation. Journal of Speech and Hearing Research, 33, 761-775.

Olson, H.F. (1967). Music, Physics and Engineering (2nd Ed.). New York: Dover Publications

Rock, A.M.L., Trainor, L.J., & Addison, T.L. (1999). Distinctive messages in infant-directed lullabies and playsongs. Developmental Psychology, 35, 527-534.

Seaver, E.J., Dalston, R.M., Leeper, H.A., & Adams, L.E. (1991). A study of nasometric values for normal nasal resonance. Journal of Speech and Hearing Research, 34, 67-73.

Scherer, K.R. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner & A. Manstead (Eds.), Handbook of Social Psychophysiology. John Wiley & Sons Ltd.

Sundberg, J. (1991). The Science of Musical Sounds. San Diego: Academic Press.

Trainor, L.J., Austin, C.M., Desjardins, R.N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? Psychological Science, 11, 188-195

Trainor, L.J., Clark, E.D., Huntley, A., & Adams, B.A. (1997). The acoustic basis of preferences for infant-directed singing. Infant Behavior and Development, 20, 383-396.

Trainor, L.J., & Trehub, S.E. (1992). The development of referential meaning in music. Music Perception, 9, 455-470.

Trehub, S.E., Endman, M.W., & Thorpe, L.A. (1990). Infants' perception of timbre: Classification of complex tones by spectral structure. Journal of Experimental Child Psychology, 49, 300-313.

Walker-Andrews, A.S., & Grolnick, W. (1983). Discrimination of vocal expression by young infants. Infant Behavior and Development, 6, 491-498.

Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of instrumental timbre. Scandanavian Journal of Psychology, 13, 228-240.

Appendix A.

Here we determine the intensity rove needed to mask local intensity cues that could potentially be used to discriminate spectral slope.

The general equation used to generate the tones was:

$$dB_i = b \log_2 i$$

where *i* is the harmonic number (from 1 to 5), *b* is the spectral slope value in dB/octave (from –16 to +16 dB/octave) and dB_i is the relative amplitude of harmonic *i* in dB. This equation was applied to each component in the complex to give the relationship in dB/octave between each successive harmonic in the complex. The tones were also equated for intensity, such that the total intensity of the complex, I_T across the 5 harmonics, was kept constant at I_T =1.00. This ensured that there would be no clipping of any of the waveforms and that the total energy for the complex was the same for each tone. To determine the overall intensity in dB for the complex, the intensity for each component in the complex needed to be calculated and then summed across the 5 harmonics. An intercept value, *a*, was added to the general equation, such that for any slope value, the total intensity could be kept constant.

$$I_T = \sum 10^{dBi/10}, \text{ where } dB_i = a + b \log_2 i$$

The equation now becomes: $I_T = \sum 10^{(a+ b \log i) / 10}$

Solving for *a*, the equation becomes:

$$a = 10 (\log_{10} I_T - \log_{10} \sum 10^{(b \log i) / 10})$$

Setting $I_T = 1.0$, the equation becomes:

$$a = -10 \log_{10} \Sigma\, 10^{(b \log i) / 10}$$

Due to the nature of the stimuli, local intensity cues could potentially be used as a basis for discrimination. The changes in spectral slope will change the absolute intensity of each of the components in the complex, and therefore local intensity cues, rather than the overall spectral envelope, could theoretically be used to detect the changes in complexes with differing spectral slopes. To minimize the use of local intensity cues, a roving intensity level needs to be introduced. To determine how much of an intensity variation was necessary, the intensity differences between the corresponding harmonics in the standard and comparison stimuli were calculated for every pair of tones that would be tested, and the maximum difference determined for each pair of tones. This difference was added to the overall intensity of the complex with the lesser slope (see A: Figure 1). The overall intensities of the two tones were then calculated to determine the intensity rove needed to mask the local intensity cues at each harmonic. For stimuli with negative spectral slopes, the harmonics that change the most in absolute intensity are the highest harmonics, and for positive spectral slopes, lowest harmonics change the most. The complexes with positive slopes differ the most at the first harmonic, while those with negative slopes are most different at the fifth harmonic. Thus, the following calculations use the first harmonic intensity values for the positive slopes and the fifth (last) harmonic intensity values for the negative slopes.

Using the general equation $dB_i = a + b \log_2 i$, the relative amplitude of the fifth (negative slopes) or first (positive slopes) harmonic, in decibels, was calculated for the standard complex with slope *b'*. The analogous harmonic of the comparison complex with slope *b"* slope was set to be equal to the intensity of the standard,

$$dB_{ib'} = dB_{ib''}$$

where *b'* is the standard slope and *b"* is the comparison slope and *i* is either harmonic number 1 or 5. Substituting the new value for $dB_{ib''}$ into Equation 1, with slope *b"* gives

$$dB_i = a + b'' \log_2 i.$$

Solving for *a*, the equation becomes:

$$a = dB_i - b'' \log_2 i.$$

This equation can now be used to solve for the new value of the harmonics with slope *b"*.

$$dB_i = (dB_i - b'' \log_2 i) + b'' \log_2 i_{b''}$$

where $i_{b''}$ is harmonic 1 to 5 in the comparison complex. This equation gives the relative amplitudes of each of the harmonics in the comparison complex. To determine the new total intensity of the complex, the component intensities must be summed.

$$I_{Tb''} = \sum 10^{\,(dBib'' / 10)}$$

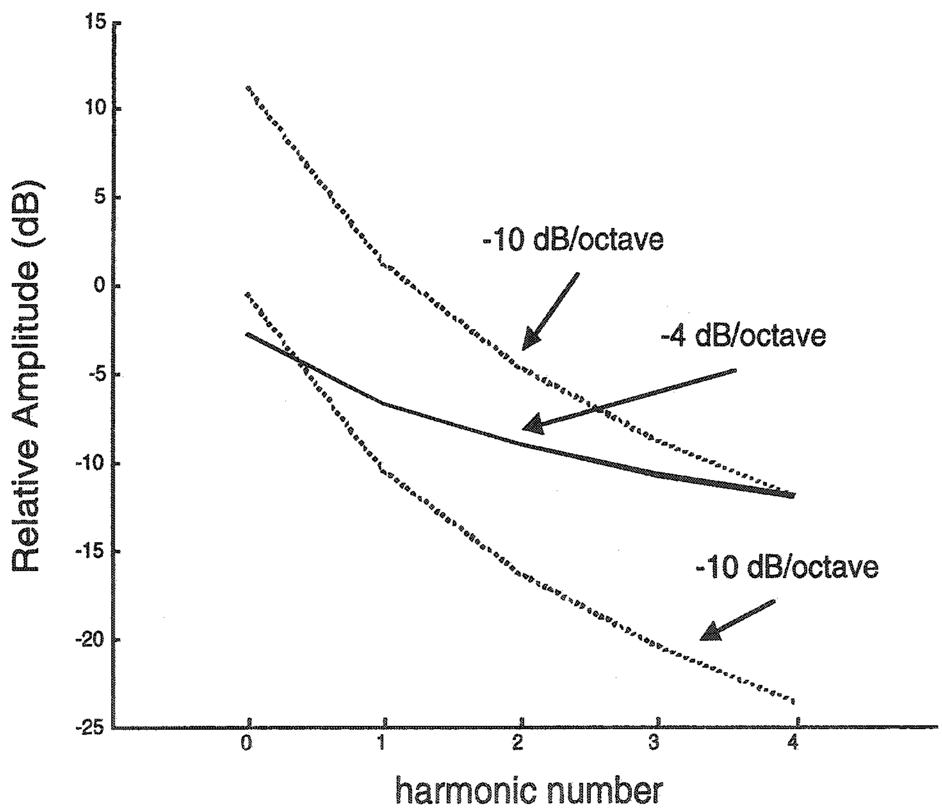
where $I_{Tb''}$ is the total intensity of the comparison complex. To determine the change in relative amplitude between the 2 complexes, which represents the intensity rove needed, the following equation was applied,

$$dB_{change} = 10 \log_{10} (I_{Tb''} / I_{Tb'})$$

where $I_{Tb'}$ is the total intensity of the standard complex, and dB_{change} is the change in relative amplitude of the between the standard and the comparison stimuli. Since the standard complex has total intensity equal to 1.0:

Setting $I_{Tb'} = 1.0$, the equation becomes:

$$dB_{change} = 10 \log_{10} (I_{Tb''})$$



A: Figure 1. Intensity differences between the –10 dB/octave (dashed lines) and –4 dB/octave (solid line) spectral slope pair. It can be seen that for negative spectral slopes, the greatest intensity change occurs at the last (fifth) harmonic. The two –10 dB/octave slopes are 11.70 dB apart. Note that the lines would appear linear if using a log scale.

CHAPTER 3: INTRODUCTION

The results from Chapter 2 demonstrate that that 8-month-old infants are able to use a global property of the spectral envelope, the spectral slope, as a means for processing and organizing auditory information, and more importantly, establish that 8-month-old infants show enhanced discrimination for sounds with spectral slope in a limited range around –10 and –4 dB/octave. This is significant because the majority of meaningful sounds in the real world have spectral slopes in this moderately negative range, and the results suggest that the infant auditory system is tuned to sounds within this limited range.

Moderately negative spectral slopes are most common to speech and music. Thus, an attentional bias for sounds with moderately negative spectral slopes may be advantageous from an infant's perspective because it would serve to attract infants' attention to sounds in the environment that are most likely to give them important information. The next chapter will examine this question by looking at attentional preference in 5- and 8-month-old infants to determine if infants show an attentional bias for sounds with environmentally common, moderately negative spectral slopes over sounds with environmentally rare, positive spectral slopes. Five-month-old infants provide an interesting contrast to 8-month-olds with respect to auditory and linguistic experience. The production of phonetic sounds begins around 5 months of age and by about 8 months, infants produce more sounds consistent with their native language than sounds found in other languages (deBoysson-Bardies, et al., 1980), suggesting that listening to one's own vocal productions may help "tune in" the infant to native language speech sounds. Thus, differences in attentional preferences to spectral slope may be seen between these age groups, as spectral slope is a potentially important cue to speech perception. The development of spectral slope sensitivity in younger 5-month-old infants, compared to 8-month-old infants, will also be examined.

The Development of Timbre Perception: Changes in Spectral Slope Perception Between 5 and 8 Months of Age

Christine D. Tsang, Laurel J. Trainor, & Joanne Leuzzi

McMaster University

Running head: Infant preferences for spectral slope

Key words: infant, auditory, preference, timbre

Address correspondence to:

Christine D. Tsang
Department of Psychology
McMaster University
Hamilton, ON, CANADA
L8S 4K1
905-525-9149 ext. 27114
fax: 905-529-6225
email: tsangcd@mcmaster.ca

Acknowledgements

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to LJT. We thank John R. Platt for his assistance in stimulus generation, Alan Rosenthal with assistance in programming, and Terri Lewis for comments on an earlier draft.

Submitted to *Developmental Science*

Abstract

Spectral slope, the linear component of the spectral envelope, affects the perception of timbre. Using a looking time procedure, we found that 8-month-old infants (Experiment 1) but not 5-month-old infants (Experiment 2) show an attentional preference for tones with –10 dB/octave slopes over tones with environmentally-rare +10 dB/octave slopes. Using a go/no-go head-turn procedure, we found that 5-month-old females discriminated tones with negative spectral slopes from tones with positive spectral slopes, but 5-month-old males did not (Experiment 3). Together, the findings show that sensitivity to spectral slope improves between 5 and 8 months of age, that sensitivity may occur earlier for females than for males, and that an attentional preference for tones with environmentally-important spectral slopes emerges by 8 months of age.

The Development of Timbre Perception: Changes in Spectral Slope Perception between 5 and 8 Months of Age

Our auditory system allows us to identify objects and people by the sounds they emit, to make social judgments about people by the quality of their voice (e.g., Bloom, Moore-Schoenmakers, & Masataka., 1999a), to identify the emotion in people's voices (e.g., Scherer, 1996; Murray & Arnott, 1993), and to discriminate the speech phonemes that enable us to understand language (e.g., Plomp, Pols, & Van de Geer, 1967). In order to understand how children come to be able to do all of these things, we are examining the development of sound quality, or *timbre*, perception. In this paper we focus on listening biases for particular timbres that may be important in early language and musical acquisition.

The primary building blocks of auditory perception are based on the acoustic features of sound. Specifically, frequency is related to the sensation of pitch, intensity is related to the sensation of loudness, and duration is related to the sensation of length. The situation is more complicated than this, however, because sounds in the world are typically made up of many frequency components. For example, a person's voice contains energy at the fundamental frequency (which is typically heard as the pitch of the voice), and also at integer multiples of that frequency. Thus, a voice producing a pitch of 100 Hz will have energy at 100 Hz, 200 Hz, 300 Hz, 400 Hz, 500 Hz, etc. Furthermore, each frequency will have a different intensity, and the temporal onsets of each frequency component may be different. The frequencies present, their energy, and their onset characteristics will all contribute to the perceived *timbre* or quality of the sound. Thus, timbre is multi-dimensional and is defined negatively as that aspect of auditory sensation by which a listener can differentiate between two sounds that are similar in pitch, overall intensity, and duration[1].

One of the main determinants of timbre is the spectral envelope (the curve connecting the points representing the intensities of the frequency components in a tonal complex) (Wedin & Goude, 1972; deBruijn, 1978). For example, different vowels have different spectral envelopes because they have different formant frequencies (resonances in the vocal tract that cause certain frequencies to be amplified in intensity). These resulting timbre differences allow us to discriminate vowels. Adult listeners are very good at detecting small changes in the spectral envelope (Green, 1983).

The linear component of the spectral envelope is called the *spectral slope* and is usually expressed in terms of decibel change per octave of frequency increase. For example, a sound with a spectral slope of –10 dB/octave has a spectrum that decreases by 10 dB for each double of frequency. Most meaningful sounds in the natural world (e.g., speech and music) have moderately negative spectral slopes (–4 to –12 dB/octave, Sundberg, 1991, p. 118; Hall, 1980, p.206), and thus lower frequency components have greater intensities than higher frequency components. Spectral slope is a cue for the identification of people (e.g., female voices have flatter spectral slopes, or more representation of high frequency components than male voices, Klatt & Klatt, 1990), the identification of objects (e.g., piano tones have flatter spectral slopes than flute tones, Olson, 1967), the identification of speech sounds (e.g., vowel sounds can be altered by

---

[1] ANSI, 1978, Psychoacoustic Terminology, s3.20

changing the degree of spectral slope, Lea & Summerfield, 1994), the identification of vocal expressions of emotion (e.g., flatter spectral slopes have been found for voices expressing fear and rage than those expressing happiness, Scherer, 1989), and is important in social interaction (speech with a flatter spectral slope is perceived as "whining" and nasal, leading to attributions of weakness, incompetence, and low intelligence, Bloom, Moore-Schoemakers, & Mastaka, 1999a; Bloom, Zajac & Titus, 1999b; Masataka & Bloom, 1994; Laver, 1980).

When does sensitivity to timbre develop? We suspect that infants are sensitive to some aspects of timbre very early in life, although much of the evidence is indirect. For example, 2-day-old infants recognize their mother's voice (DeCasper & Fifer, 1980) and 2-month-olds can discriminate male and female voices (Miller, 1983). Timbre is certainly a strong cue in these situations, but there are also other cues, such as talker-specific intonation patterns. Timbre is also a distinguishing feature of infant-directed singing, in that adult listeners rate infant-directed singing as being rendered in a more “loving tone of voice” than non-infant-directed singing, and infants listen longer the more loving the tone of voice (Trainor, 1996). Spectral slope may contribute to this preference. First, infant-directed singing has a more negative spectral slope than non-infant-directed singing (Trainor, Clark, Huntley, & Adams, 1997). Second, the expression of happiness in adult speech is also accompanied by an increased representation of low frequency energy (Scherer, 1989). On the production side of infants’ communication, the timbre of infants' vocalizations in the first months of life affects adults' rating of those infants. Specifically, infants who produce sounds that are less negatively sloped (and therefore more nasal) are rated as less cute and less intelligent than infants producing more negatively-sloped sounds (Bloom et al., 1999a, b). The strongest direct evidence for timbre discrimination in very young infants comes from studies showing that infants as young as 2 or 3 months are able to discriminate between vowel categories despite variability in pitch and talker (Marean, Werner, & Kuhl, 1992; Kuhl, 1979).

The direct evidence for timbre discrimination in the second half-year of life is stronger. By 7- to 8-months of age, infants are able to discriminate sounds that differ only in the shape of the spectral envelope (Clarkson, Clifton, & Perris, 1988; Trehub, Endman, & Thorpe, 1990). Moreover, 8-month-olds (Chapter 2), like adults (Chapter 4), show optimal spectral slope discrimination for tones with spectral slopes in the range that is most common in speech and music. Specifically, 8-month-olds are best at discriminating sounds differing in spectral slope when the slopes are in the moderately negative range (–4 to –10 dB/octave), and performance is worse for sounds either with highly negative or positive spectral slopes.

The focus of the present paper is to determine whether this enhanced discrimination is associated with an attentional bias to listen to sounds with moderately negative spectral slopes over those with positive spectral slopes, and at what age such a bias might emerge. An attentional bias for sounds with moderately negative spectral slopes would serve the very important function of attracting infants' attention to sounds in the environment that are most likely to give them important information, in particular, speech and music.

Infants often show attentional biases for stimuli that are prevalent and/or important in the natural world. For example, newborn infants spend more time looking at some face-like stimuli than at non-face-like stimuli (e.g., Morton & Johnson, 1991; Mondloch, et al., 1999). In the auditory domain, infants prefer to listen to voices, speech, and music over other auditory stimuli (Butterfield & Siperstein, 1972; for a review, see Standley & Madson, 1990). Voices, speech and music are all prevalent in the human infant's environment and contain important information for language and social communication. Furthermore, people talk and sing to infants with high pitch, and infants prefer higher- over lower-pitched speech (Patterson, Muir, & Hains, 1997) and music (Trainor & Zacharias, 1998). Finally, infants as young as 2 months of age prefer consonant (pleasant-sounding to adults) over dissonant (unpleasant-sounding to adults) musical chords and, although both are present, consonance is more common than dissonance in music (Trainor, Tsang, & Cheung, in press).

In the present paper, we examined whether the acoustic feature of spectral slope biases infants' listening behaviour. To ensure that we were testing this feature rather than familiarity for speech or music per se, we used unfamiliar computer generated steady-state tones that had spectral slopes of either –10 dB/octave or +10 dB/octave. In Experiment 1, we tested whether 8-month-olds prefer tones with the negative over the positive spectral slope. In subsequent experiments, we examined the emergence of this listening bias in younger infants.

Experiment 1

Method

Participants. Twenty (10 male, 10 female) normal, full-term 8-month-old infants (*M* age= 36.49 weeks, range = 35 – 39 weeks) participated. All infants were born within 2 weeks of term, weighed at least 2500 g at birth, had no history of ear infections or a family history of hearing impairment, and were healthy at the time of testing. All infants tested completed the entire procedure (none were excluded).

Stimuli. Complex tones were generated by a Symbolic Sound Corporation, Kyma Sound Design Workstation, version 4.5. Stimuli were generated with 16-bit precision at a sampling rate of 44.1 kHz and recorded as sound files for stimulus presentation. All tones had a fundamental frequency of 200 Hz (which is in the range of female voices) and were 1000 ms in duration, including linear 20-ms rise and fall times. Each tone consisted of five harmonics in cosine phase. Two tones were used, one tone with a spectral slope of –10 dB/octave and one with a spectral slope of +10 dB/octave. Each tone was presented at six different intensity levels (65, 67, 69, 71, 73, 75 dB(A)) over a noise floor of 24 dB(A) in order to minimize infants' use of local intensity cues, that is, the possibility of listening to changes in the intensity of only one harmonic rather than changes in the spectral slope of the entire sound (see Chapter 2). On negative spectral slope trials, the –10 dB/octave tone repeated continuously with 250 ms between tones, and the intensity on each repetition was chosen randomly from the set six intensity levels. Similarly, on positive spectral slope trials, the +10 dB/octave tone repeated continuously with 250 ms between tones, and the intensity on each repetition was chosen randomly from the set of six intensity levels.

Apparatus. Testing was conducted in a sound-attenuating chamber (Industrial Acoustics Co.). A Power Macintosh 7200/180 computer ran the experiment. The sound stimuli were passed through a Denon amplifier (PMA-480) to a pair of loudspeakers (GSI) located inside the chamber. The parent holding the infant sat in a chair arranged such that one loudspeaker was located on the infant's right and the other located on the infant's left. Under each loudspeaker was a compartment containing lights and a mechanical toy. The compartments had smoked plexiglass fronts such that when a light was illuminated inside the compartment, the toy became visible. A custom-built interface box connected the button box (used by the experimenter to signal to the computer) and lights to the computer. The experimenter was seated facing the infant.

Procedure. Infants were tested individually in a visually-based preference procedure. Trials of one spectral slope alternated with trials of a second spectral slope, with the length of each trial dependent on the infant's head-turning behavior. The experimenter and the parent listened to masking music through headphones so as to be unaware of what the infant was hearing. For a given infant, trials with the negative spectral slope (–10 dB/octave) were always presented on one side (right or left) and trials with the positive spectral slope (+10 dB/octave) were always presented on the other. Half of the infants tested received trials of the negative spectral slope on the left, and trials of the positive spectral slope on the right; the other half received the reverse. In each case, the side of presentation alternated between trials, with the initial side (right or left) and initial spectral slope (+10, –10 dB/octave) crossed and randomized across subjects. The experimenter was not aware of which spectral slope was being presented on which side for each subject.

The sound stimuli and lights were controlled by the computer. When the infant was attentive and facing forward, the experimenter pressed one button on the button box (held out of the infant's view) to initiate a trial. This caused the light on one side to begin flashing (400 ms on, 400 ms off), illuminating a toy in the box under the speaker. When the infant turned to look at the light and the toy, the experimenter pressed a second button that resulted in the presentation of the appropriate sound stimulus for that side. The light remained on, but ceased to flash during the sound presentation. The experimenter held the button down as long as the infant continued to look at that side. The sound presentation continued until the infant looked away for at least 2 s, and the looking time was recorded by the computer. The light (which could be seen by the experimenter) and sound (which could not be heard by the experimenter) were extinguished at the end of the trial. Testing ended when the infant completed 18 trials (9 trials with the negative spectral slope, 9 trials with the positive spectral slope).

Results and Discussion

The first two trials (one with positive spectral slope and one with negative spectral slope) were eliminated from the analysis on the rationale that at the beginning of the testing session the infant was not aware that he/she controlled the length of stimulus presentation. Because of high between subject variation in total looking time (mean looking time per trial = 4.19 seconds, range = 1.96 seconds to 8.29 seconds), analyses on the remaining 16 trials were conducted on the proportion of total looking time in seconds to the negative spectral slope. Thus, the sum of looking times across all negative spectral slope trials was divided by the sum of looking times across all trials.

Before testing whether infants showed significant preferences for negative spectral slopes, we wished to establish whether there were effects of which spectral slope was presented on which side, which spectral slope was heard first, and sex of the infants. It was not possible to do a single ANOVA with all of these variables as they were not completely crossed. However, an ANOVA was conducted with proportion looking time to the negative spectral slope as the dependent variable and side of first presentation (right, left) and spectral slope heard first (positive, negative) as between subjects variables. No significant effects were found. A further ANOVA with sex as a between subjects factor showed no significant effects of sex. Thus, we collapsed across side of first presentation, spectral slope heard first, and sex for the subsequent analysis. A one-sample *t*-test with an expected value of .5 under the null hypothesis revealed no significant difference in the amount of time spent looking to the tone with negative spectral slope compared to the positive spectral slope.

Because infants often habituate over repeated stimulus presentations (see Slater, Carrick, Bell, & Roberts, 1999 for a review), a second analysis was conducted using only the first eight trials (trials 3 to 10). Again, it was found that there were no effects of side of presentation, slope heard first or sex. However, a one-sample *t*-test, with an expected value of .5 under the null hypothesis, showed that infants looked significantly longer in order to hear the tones with negative spectral slope $t(19) = 2.28$, $p < 0.03$ (Figure 1). No significant effects were found for the second (last) eight trials.

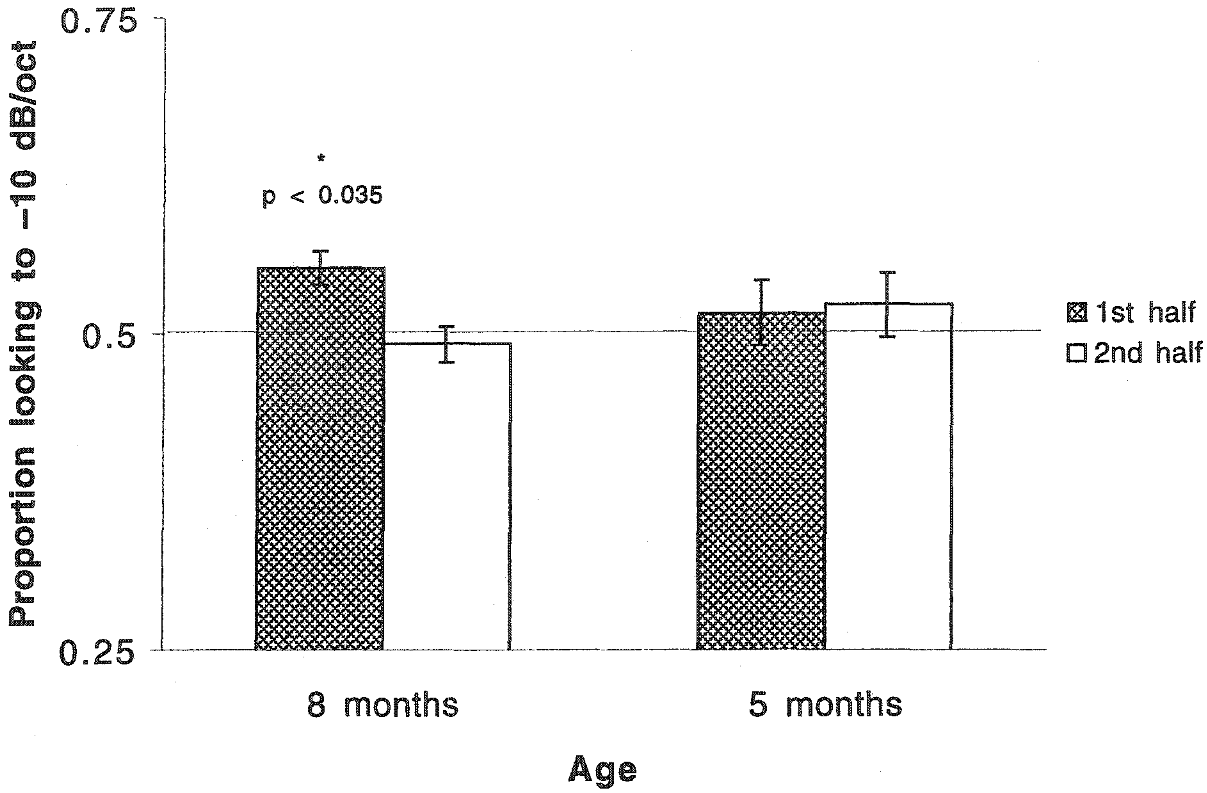


Figure 1. Proportion looking times to the tone with −10 dB/octave spectral slope in the first half of the testing session (hatched bars) and the second half of the testing session (white bars) in 8-month-olds (Experiment 1) and 5-month-olds (Experiment 2). Error bars represent within-subject confidence intervals (Loftus & Masson, 1994).

The results indicate that 8-month-old infants show an attentional preference for sounds with a negative spectral slope over those with a positive spectral slope. Thus, not only are 8-month-olds best able to discriminate the moderately negative spectral slope differences that matter in speech and music (Chapter 2), but they listen longer to unfamiliar sounds with spectral slopes in this range. Such a listening bias will serve to attract infants' attention to speech and music.

In the next experiment, we asked how early such a listening bias is present. Specifically, we examined the attentional preferences of 5-month-old infants to determine if they also prefer to listen to sounds with negative over positive spectral slopes.

Experiment 2

Method

Participants. Forty (21 male, 19 female) normal, full-term 5-month-old infants (*M* age = 23.8 weeks, range = 21.7 weeks – 25.5. weeks) participated. All infants were born within 2 weeks of term, weighed at least 2500 g at birth, had no history of ear infections or family history of hearing impairment, and were healthy at the time of testing. A further 3 5-month-old infants were excluded for colds and/or crying.

Stimuli, Apparatus, & Procedure. Identical to those of Experiment 1.

Results and Discussion

As in Experiment 1, all infants in the final sample completed 18 trials, the first two of which were eliminated from the analyses. Again, analyses were conducted on the proportion of looking time to the negative spectral slope during the first and second halves of the test session. The same initial ANOVAs were conducted as for the 8-month-olds in Experiment 1. These revealed no significant main effects or interactions involving side of first presentation (right, left), spectral slope heard first (negative, positive), or sex (male, female). However, the subsequent *t*-tests revealed no significant listening preferences for negative or positive spectral slopes in either the first or the second halves of the test sessions (Figure 1).

The results of Experiment 2 suggest that, unlike 8-month-old infants, 5-month-olds do not show an attentional preference based on spectral slope. It is unlikely that this lack of preference is caused by problems using the head-turn preference methodology with 5-month-old infants. By 4 months of age, infants are able to perform reliable short-latency head-turns (Werker, Polka, & Pegg, 1997). Moreover, the head-turn preference procedure has been used successfully with infants as young as 4.5 months to determine both intrinsic preferences and discriminative abilities (for a review, see Kemler-Nelson, Jusczyk, Mandel, Myers, Turk, & Gerken, 1995). One possibility for the lack of preference is that 5-month-olds are unable to even discriminate the stimuli. Thus, in Experiment 3, we used a conditioned head turn procedure (as in Chapter 2) to test whether 5-month-old infants can discriminate tones with spectral slopes of +10 and –10 dB/octave.

Experiment 3

Participants. Sixteen (8 male, 8 female) normal, full-term 5-month-old infants (*M* age = 21.56 weeks, range = 20.4 – 24.3 weeks) participated. All infants were born within 2 weeks of term, weighed at least 2500 g at birth, had no history of ear infections or family history of hearing impairment, and were healthy at the time of testing. An additional 8 infants (7 males, 1 female) were eliminated from the final sample because of failure to complete both phases of the procedure (see Procedure).

Stimuli. The stimuli were identical to those described in Experiments 1 and 2.

Apparatus. The apparatus was the same as in Experiments 1 and 2, with the following exceptions. The sound stimuli were passed through the amplifier to a single loudspeaker located inside the chamber on the infant's left. Under the loudspeaker was a compartment containing computer-controlled lights and animated, mechanical toys which were used for reinforcement (see Procedure).

Procedure. Infants were tested individually in a go/no-go conditioned head-turn procedure. The experimenter and the parent listened to masking music through headphones so as to be unaware of what the infant was hearing. During the experiment, one of the spectral slopes (either +10 or –10 dB/octave), the standard spectral slope, repeated continuously from a loudspeaker on the infant's left, with the intensity level chosen randomly on each repetition. The interval between the tones was 1000 ms. When the infant was attentive and facing the experimenter, the experimenter pressed a button on the button box (held out of the infant's sight) to indicate to the computer to initiate a trial. There were two types of trials: control trials, in which a tone with the standard spectral slope (either +10 or –10 dB/octave) was presented, and change trials, in which a tone with the other spectral slope was presented. The experimenter indicated infant head-turns to the computer by pressing a second button on the button box. If the infant made a 45° head turn toward the loudspeaker within 3000 ms of the beginning of a change trial, a toy in one of the compartments under the loudspeaker was activated and lit up by the computer for 3000 ms as a reinforcer. Head turns at other times, and those that were less than 45°, were not reinforced. The computer kept track of any head turns that occurred within a 3000 ms window beginning at the onset of the tone on change trials (hits) as well as the onset of the tone on control trials (false alarms) within a 3000 ms window to provide an index of the false alarm rate. Trials were presented in a quasi-random order for each subject, with the constraint that no more than two control trials occurred sequentially.

For half of the infants tested, tones of –10 dB/octave were presented as the standard and tones with +10 dB/octave were presented during change trials. For the other half of the infants, this was reversed. There were two experimental stages, Phase 1 and Phase 2. During Phase 1, only change trials were presented and intensity was held constant at 71 dB(A) to make the task easier. Demonstration trials, in which the change slope was presented if the infant failed to turn on several trials in a row, were used in order to show the infant that head-turning to a change tone would be rewarded. A criterion was set at 4 correct trials in a row within 20 trials. If the infant failed to reach criterion, the session was terminated. If the infant reached criterion, Phase 2 began. During this phase, both the standard and change tones were presented with the full range of intensity levels (65 to 73 dB(A)), with the intensity on each stimulus repetition chosen randomly. There were no demonstration trials during Phase 2. Each infant completed 24 trials in Phase 2: 12 change trials and 12 control trials.

Results and Discussion

Sixteen (8 males and 8 females) out of twenty-four infants tested reached criterion in Phase 1 and completed Phase 2. For each infant completing Phase 2, a d' value was calculated using the proportions of hits and false alarms. An ANOVA with condition (–10/+10 dB/octave control/change slopes, +10/–10 dB/octave control/change slopes) and sex(male, female) as factors and d' as the dependent variable revealed only a main effect of sex, $F(1, 12) = 11.50, p < .005$. Female infants performed significantly above chance levels (expected value of d'=0 under the null hypothesis), $t(7) = 3.331, p < 0.01$ (mean $d' = 0.54$), but male infants did not, $t(7) = 0.537, p < 0.61$ (mean $d' = -0.10$) (Figure 2).

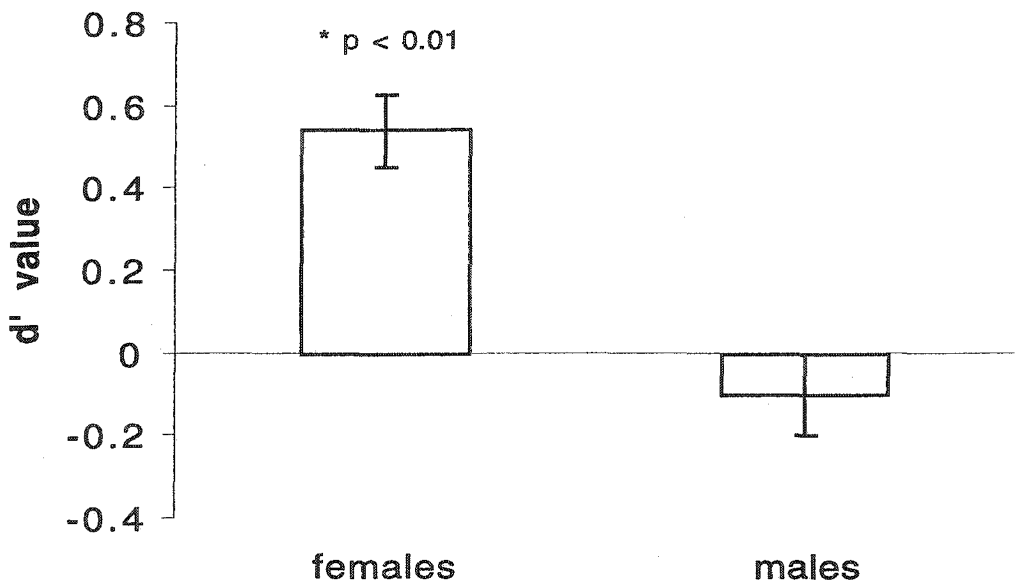


Figure 2. d' values for female infants compared to male infants in Experiment 3. Chance performance has an expected value of 0. Error bars represent the standard error of the mean.

Previous studies indicate that 7- to 8-month-olds of both sexes can discriminate sounds differing only in spectral slope (Clarkson, 1996; Chapter 2). It is unlikely that the present negative results with male 5-month-olds are due to procedural problems, as female infants were able to perform above chance levels, and the conditioned head-turn procedure has previously been used successfully with infants as young as 5 months (e.g., Werker, Polka & Pegg, 1997; Kemler-Nelson, et al., 1995). Thus, we conclude that there is an improvement in spectral slope discrimination between 5 and 8 months of age, and that sensitivity to spectral slope differences develops earlier in females than in males. In general, females tend to develop auditory skills earlier than males, especially in language related tasks (Darley & Winitz, 1961; Rome-Flanders & Cronk, 1995), and at 3 months, female infants show more adult-like auditory evoked potentials to verbal stimuli than do male infants (Shucard, Shucard, Cummings & Campos, 1981).

However, the ability of female infants to discriminate sounds with a positive spectral slope from sounds with a negative spectral slope does not appear to result immediately in an attentional bias to listen to negatively over positively sloped sounds at 5 months, as no preference was seen in Experiment 2. There was no significant difference in preference between males and females on the preference task of Experiment 2. However, just to make sure that there were no trends for a sex difference in Experiment 2, the data were analyzed separately for males and females. Neither group showed a significant preference for either spectral slope in the first half (males, $p = 0.6$, females, $p = 0.7$) or the second half (males, $p = 0.62$, females, $p = 0.55$). This suggests that although 5-month-old female infants have begun to develop sensitivity to spectral slope, they have not yet developed the biased listening seen in 8-month-olds.

General Discussion

The results show a clear developmental progression in sensitivity to one important determinant of timbre, namely, spectral slope. Sensitivity to spectral slope appears to emerge some time around 5 months of age, as females, but not males, of this age can discriminate an unfamiliar sound with a moderately negative spectral slope from an unfamiliar sound with a moderately positive spectral slope (Experiment 3). However, there is no evidence of an attentional bias for sounds with negative spectral slopes at this age (Experiment 2). By 8 months, however, both sexes not only discriminate positive from negative spectral slopes (Clarkson, 1996), but show enhanced fine discrimination for moderately negative spectral slopes (Chapter 2) and an attentional listening bias for unfamiliar sounds with negative spectral slopes (Experiment 1, this paper). Thus, by 8 months, infants are biased to attend to new sounds with negative over positive spectral slopes, a bias that will direct their attention more to speech and music over other stimuli.

Attentional biases and discrimination abilities are caused by some combination of innate and learned processes. For example, infant preferences for high pitch may partly result from an innately-based tendency to associate high pitch with friendliness and non-aggression, an association that is seen across animal species (e.g., Morton, 1977). However, experience may also be partly responsible for infants' preferences for high pitch because both infant-directed speech and infant-directed singing are higher in pitch than their adult-directed counterparts (Fernald & Kuhl, 1987; Trainor et al., 1997; Trainor & Zacharias, 1998).

Along the same lines, infant preferences and enhanced discrimination abilities at 8 months for negative spectral slopes may be caused by innate biases that evolved to help infants attend to meaningful environmental sounds. Alternatively, they may be caused by the effects of experience, because by this age infants have had a lot of experience with negatively-sloped speech and musical sounds. Thus, a listening bias for sounds with negative spectral slopes may cause infants to attend more to speech and music than to other auditory sounds, or the fact that infants attend to speech and music may cause them to prefer and show greater sensitivity to sounds with negative spectral slopes.

The data presented here cannot definitively distinguish between these two possibilities. However, an argument in favour of the hypothesis that preferences follow discrimination can be made. First, the discrimination of spectral slope occurs earlier than a preference for negative spectral slopes. Second, the changes in spectral slope processing seen between 5 and 8 months of age parallel important changes in speech development across these ages. Although language development occurs throughout the first year of life and beyond, the production of phonetic sounds generally does not occur before 5 or 6 months of age (for a review, see deBoysson-Bardies, Viham, Roug-Hillichius, Durand, Landberg, & Arao, 1992). One hypothesis related to the development of speech perception suggests that infants may learn to discriminate and categorize phonemes in their native language partly by listening to their own vocal productions (Aslin & Pisoni, 1980). Thus, one difference between 5- and 8-month-olds is that 8-month-olds have had much more experience listening to their own vocalic speech productions and have become sensitive to the phonemic categories of their native language (Werker & Tees, 1983). Third, infants are likely rewarded for producing sounds with more negative spectral slopes. Masataka and Bloom (1994) showed that mothers were less responsive to infant vocalizations with more positive spectral slopes, which were also rated by adults as having higher nasality. In sum, the spectral slope processing differences seen between 5 and 8 months may reflect experience with vocal production and maternal rewards for producing more negatively-sloped vocal sounds.

The sex differences seen in Experiment 3 also support the notion that linguistic experience may underlie the developmental changes observed between 5 and 8 months. Females generally show superior performance at an earlier age in many language-related tasks (e.g., verbal fluency, articulation) beginning in early infancy (Darley & Winitz, 1961; Rome-Flanders & Cronk, 1995). This earlier linguistic development and earlier vocal production experience in females may contribute to the sex differences in spectral slope discrimination observed in Experiment 3.

Whatever the underlying developmental mechanism, we have shown a clear progression in spectral slope perception between 5 and 8 months of age. Spectral slope gives us information about people and objects in the world, as it varies systematically across different musical instruments, different vowels, different talkers, and different vocal expressions of emotion. Infants' increasing sensitivity to this dimension of timbre across the second half-year of life is one reflection of their increasing awareness of the physical and social worlds.

References

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G.H. Yeni-Komshian, J.F. Kavanagh, & C.A. Ferguson (Eds.) Child Phonology, Volume 2 Perception (pp. 67-96). New York, NY: Academic Press, Inc.

Bloom, K., Moore-Schoenmakers, K., & Masataka, N. (1999a). Nasality of infant vocalizations determines gender bias in adult favorability ratings. Journal of Nonverbal Behavior, 23, 219-235.

Bloom, K., Zajac, D. J., & Titus, J. (1999b). The influence of nasality of voice on sex-stereotyped perceptions. Journal of Nonverbal Behavior, 23, 271-281.

Butterfield, E. C., & Siperstien, G. N. (1972). Influence of contingent auditory stimulation upon nonnutritional suckle. In J. Bosma (Ed.) Oral Sensation and Perception: The Mouth of the Infant (p.313-333). Springfield, IL: Charles C. Thomas.

Clarkson, M. G. (1996). Infants' intensity discrimination: Spectral profiles. Infant Behavior and Development, 19, 181-190.

Clarkson, M. G., Clifton, R. K., & Perris, E. E. (1988). Infant timbre perception: Discrimination of spectral envelopes. Perception and Psychophysics, 16, 564-570.

Darley, F. L., & Winitz, H. (1961). Age of first word: Review of research. Journal of Speech and Hearing Disorders, 26, 272-290.

DeBoysson-Bardies, D., Viham, M. M., Roug-Hillichius, L., Durand, C., Landberg, I., & Arao, F. (1992). Material evidence of infant selection from the target language: a cross-linguistic phonetic study. In C. A.. Ferguson, L. Menn, C. Stoel-Gammon (Eds.) *Phonetic Development: Models, Research, Implications*, Timonium, Maryland: York Press, Inc. pp. 369-391.

DeBruijn, A. (1978). Timbre classification of complex tones. Acustica, 40, 108-114.

DeCasper, A. J., Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. Science, 208, 1174-1176.

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

Green, D. M. (1983). Profile analysis. A different view of auditory intensity discrimination. American Psychologist, February, 133-142.

Hall, D. E. (1980). Musical Acoustics: An Introduction. Belmont, California: Wadsworth Publishing Co.

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

Klatt, D. H. & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America, 87, 820-857.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America, 66, 1668-1679.

Laver, J. (1980). The Phonetic Description of Voice Quality. Cambridge: Cambridge University Press.

Lea, A. P., & Summerfield, Q. (1994). Minimal spectral contrast of formant peaks for vowel recognition as a function of spectral slope. Perception and Psychophysics, 56, 379-391.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. Psychonomic Bulletin & Review, 1, 476-490.

Marean, G. C., Werner, L. A., & Kuhl, P. K. (1992). Vowel categorization by very young infants. Developmental Psychology, 28, 396-405.

Masataka, N., & Bloom, K. (1994). Acoustic properties that determine adults' preferences for 3-mont-old infant vocalizations. Infant Behavior and Development, 17, 461-464.

Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. Infant Behavior and Development, 6, 313-330.

Mondloch, C.J., Lewis, T.L., Budreau, D.R., Maurer, D., Dannemiller, J.L., Stephens, B.R., & Kleiner-Gathercoal (1999). Face perception during early infancy. Psychological Science, 10, 419-422.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. American Naturalist, 111, 855-869.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLEARN: A two-process theory of infant face recognition. Psychological Review, 98, 164-181.

Murray, I. R. & Arnott, J. L. (1993). Toward simulation of emotion in synthetic speech, A review of the literature on human vocal emotion. Journal of the Acoustical Society of America, 93, 1097-1108.

Olson, H. F. (1967). Music, Physics and Engineering (2nd Ed.). New York: Dover Publications.

Patterson, M., Muir, D. W., & Hains, S. M. J. (1997). Infant sensitivity to perturbations in adult infant-directed speech during social interactions with mother and stranger. Poster presented at the biennial meeting of the Society for Research in Child Development, Washington, D.C.

Plomp, R., Pols, L. C., & Van de Geer, J. P. (1967). Dimensional analysis of vowel spectra. Journal of the Acoustical Society of America, 41, 707-712.

Rome-Flanders, T., & Cronk, C. (1995). A longitudinal study of infant vocalizations during mother-infant games. Journal of Child Language, 22, 259-274.

Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner & A. Manstead (Eds.), Handbook of Social Psychology. John Wiley & Sons Ltd.

Shucard, J. L., Shucard, D. W., Cummins, K. R., & Campos, J. J. (1981). Auditory evoked potentials and sex-related differences in brain development. Brain and Language, 13, 91-201.

Slater, A., Carrick, R., Bell, C., & Roberts, E. (1999). Can measures of infant information processing predict later intellectual ability? In A. Slater & D. Muir (Eds.), The Blackwell Reader in Developmental Psychology, Masachussetts: Blackwell Publishing, pp. 55-64.

Standley, J. M., & Madsen, C. K. (1990). Comparison of infant preferences and responses to music, mother and other female voice. Journal of Music Therapy, 2, 54-97.

Sundberg, J. (1991). The Science of Musical Sounds. San Diego: Academic Press.

Trainor, L. J. (1996). Infant preferences for infant-directed versus non-infant-directed playsongs and lullabies. Infant Behavior and Development, 19, 83-92.

Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis of preferences for infant-directed singing. Infant Behavior and Development, 20, 383-396.

Trainor, L. J., Tsang, C. D., & Cheung, V. H. W. (in press). Preference for musical consonance in 2-month-old infants. Music Perception.

Trainor, L. J., & Zacharias, C. A. (1998). Infants prefer higher-pitched singing. Infant Behavior and Development, 21, 799-805.

Trehub, S. E., Endman, M. W., & Thorpe, L. A. (1990). Infants' perception of timbre: Classification of complex tones by spectral structure. Journal of Experimental Child Psychology, 49, 300-313.

Tsang, C. D., & Trainor, L. J. (2002). Spectral slope discrimination in infancy: Sensitivity to socially important timbres. Infant Behavior and Development, 25, 183-194.

Tsang, C. D., & Trainor, L. J. (submitted). Discrimination of Spectral Slope in Adult Listeners.

Wedin, L., & Goude, G. (1972). Dimension Analysis of the perception of instrumental timbre. Scandanavian Journal of Psychology, 13, 228-240.

Werker, J. F., Polka, J., & Pegg, J. E. (1997). The conditioned head-turn procedure as a method for testing infant speech perception. Early Development and Parenting, 6, 171-178.

Werker, J. F., & Tees, R. C. (1983). Cross language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 7, 49-63.

CHAPTER 4: INTRODUCTION

The studies reported in Chapters 2 and 3 demonstrate that there is a developmental progression in sensitivity to spectral slope during infancy. While 8-month-old infants show both optimal discrimination for environmentally common, moderately negative spectral slopes as well as an attentional preference for sounds with these characteristics, 5-month-olds show no such preference, and only female 5-month-olds show discrimination between positive and negative spectral slopes. However, in order for these results to be compared in any meaningful way to adult auditory processing, a systematic examination of adult spectral slope discrimination must be conducted.

Adult listeners have been found to be highly sensitive to global spectral characteristics of sounds. Studies using profile analysis have demonstrated that adults use spectral shape differences to make discriminations (e.g., Green, 1983; Green & Kidd, 1982; Green, Kidd, & Picardi, 1983; Speigel, Picardi, & Green, 1981; Versfeld & Houstma, 1991). Studies showing the perceptual independence of spectral slope from fine spectral characteristics of sound (e.g., number of components, location of resonances) have found that the dimension of spectral slope may play an important role in identifying perceptual constancies in sounds, such as recognizing the same voice across varying speech sounds (Li & Pastore, 1995). In the few studies that have directly examined spectral slope discrimination, listeners have been found to discriminate between moderately negative spectral slopes (Li & Pastore, 1995) and tilted from flat spectra (Bernstein & Green, 1987). However, to date, there has been no reported study systematically examining adults' thresholds across a wide range of spectral slope values. This chapter will examine adult sensitivity to spectral slope, and determine if adult listeners show enhanced sensitive to moderately negative spectral slopes, similar to 8-month-old infant listeners.

Spectral Slope Discrimination

Christine D. Tsang & Laurel J. Trainor

Department of Psychology, McMaster University

Hamilton, Ontario, Canada, L8S 4K1

Running head: Spectral slope and timbre perception

Address correspondence to:

Christine D. Tsang
Department of Psychology
McMaster University
Hamilton, Ontario, Canada
L8S 4K1
phone: (905) 525-9140 ext. 27114
fax: (905) 529-6225
e-mail: tsangcd@mcmaster.ca

Acknowledgements

This research was supported by a grant from the National Science and Engineering Council of Canada to LJT and an Ontario Graduate Scholarship to CDT. The authors would also like to thank John R. Platt for assistance with sound generation, Patrick J. Bennett, Nicole D. Anderson, and Alan J. Rosenthal for technical assistance, Joanne Leuzzi for assistance with data collection, and Dennis P. Phillips for helpful comments on an earlier draft.

Abstract

Spectral slope is a global property of the spectral envelope that contributes to the perceived timbre differences between voices and musical instruments. The ability to discriminate between spectral slopes develops early in life, and infants show optimal discrimination for negative spectral slopes in the range of speech and music (–10 to –4 dB/octave). The current study showed that (1) for adults, average 85% thresholds for spectral slope discrimination range between 1.5 and 4 dB/octave for standard spectral slopes between –18 and + 18 dB/octave, and (2) performance is best for negative spectral slopes in the range of those seen in speech and musical sounds.

I. INTRODUCTION

Timbre is defined as the perceptual difference between sounds that are equivalent in pitch, loudness and duration[1]. Despite the imprecise nature of this definition, timbre is critical for the discrimination of musical instruments, voices, and speech sounds, as well as for the segregation of simultaneous sounds and for hearing in noise. Multiple acoustic dimensions of sound contribute to the perception of timbre, making a systematic study difficult. Studies using multidimensional scaling (MDS) to examine the perceived similarity of musical instruments (Grey, 1977; Grey & Gordon, 1978; Iverson & Krumhansl, 1993; McAdams, Winsberg, Donnadieu, DeSoete, & Krimpoff, 1995; McAdams & Cunible, 1992; Miller & Carterette, 1975) and speech sounds (DeBruijn, 1978; Plomp, Pols, & Van de Geer, 1966; Pols, Van der Kamp, & Plomp, 1969) have found that the timbre of steady-state complex stimuli depends critically on overall spectral envelope characteristics such as bandwidth, cutoff frequencies, and spectral shape.

The effect of spectral shape on perception has been examined using profile analysis, whereby listeners must detect an intensity increment in a single frequency component in the spectrum of a complex tone. A change in a single frequency component affects the relative amplitudes of the frequency components in the complex (i.e., the profile or spectral shape), and there is much evidence that listeners rely on this spectral shape change rather than the local intensity increment for discrimination. First, from a subjective perspective, listeners have reported relying on perceived timbre changes to perform the task (Versfeld & Houtsma, 1991). Second, variation in the overall intensity level of the complex tones being compared has little effect on performance, with listeners able to detect a relative intensity change as small as 1-2 dB in a single frequency component in complex of over 20 components (Green, 1983; Green & Kidd, 1982; Green, Kidd, & Picardi, 1983; Spiegel, Picardi, & Green, 1981; Versfeld & Houtsma, 1991). Third, under conditions where the spectrum changes randomly, but the frequency component to which the signal is added remains fixed, listeners' thresholds for signal detection fall off dramatically when compared to conditions where the spectrum is fixed, but the single frequency component varies (Spiegel et al., 1981). Fourth, increasing the number of components in the spectrum actually improves performance in a profile analysis task, and indicates that listeners use the whole spectrum rather than only those components immediately adjacent to the incremented component (Green, et al., 1983). Together, these studies show that the human auditory system is very sensitive to small differences in spectral shape.

The linear component of the spectral envelope is the spectral slope and is generally expressed in terms of decibel change per octave of frequency increase. A tone complex with a positive spectral slope will have relatively greater energy in the high frequency components, while a tone with a negative spectral slope will have greater energy in the low frequency components. There is evidence that the perceptual processing of spectral slope is independent of that of other stimulus features. For spectral slopes between –6 and –24 dB/octave, spectral slope is perceptually separable from rise time characteristics, and even to a large extent from cut-off frequency (Campbell, 1994). Spectral slope also appears to be analyzed independently from fine spectral structure.

Using sounds with moderately negative spectral slopes, Li and Pastore (1995) found no decrement in spectral slope discrimination thresholds when the fine structure of the spectrum (i.e., number and location of resonances) varied randomly from sound to sound, in comparison to the situation where the fine structure remained constant. This suggests that the linear component of the spectral envelope (spectral slope) is extracted independently from higher-order components in the envelope. Li and Pastore (1995) linked spectral slope to source characteristics of voices, which remain relatively constant, and the fine spectral structure to filter characteristics of the vocal tract, which change from speech sound to speech sound. The ability to capture the source characteristics in the presence of variation in the filter characteristics would be useful for identifying perceptual constancies, such as recognizing the same voice across changing speech sounds. Indeed, spectral slope provides a cue to a number of important auditory abilities, such as voice recognition (male and female voices differ in spectral slope steepness, Klatt & Klatt, 1990), vocal quality identification (e.g., less negative (flatter) spectral slopes in the region of the lower formants are correlated with highly nasal voices, Beddor, 1993; Hawkins & Stevens, 1995), and vocal emotional identification (e.g., expressions of negative emotions such as fear and rage are associated with flatter spectral slopes than are expressions of positive emotions such as happiness, Scherer, 1989).

The average spectral slope of the human voice has been measured to vary between about –12 and –4 dB/octave (p. 206, Hall, 1980; p.118, Sundberg, 1991). We analyzed the spectral slope of published spectra for different musical instruments (from Olson, 1967 and Fletcher & Rossing, 1991) and found that the same range also applies to musical instruments. Perceptual studies show that adults can discriminate spectral slopes in the moderately negative region (Li & Pastore, 1995), and tilted spectra from flat spectra (Bernstein & Green, 1987). Campbell reported an informal study in which practiced listeners achieved spectral slope discrimination thresholds as low as 1 dB/octave in the negative spectral slope region (Campbell, 1994, note 3). However, we know of no study to date that has directly examined adults' thresholds for spectral slope discrimination across a wide range of spectral slope values.

The ability to discriminate spectral slope is also present very early in life. Infants can discriminate sounds with positive from sounds with negative spectral slopes (Clarkson, 1996). Furthermore, infants show optimal discrimination of spectral slopes in the moderately negative range (–10 to –4 dB/octave) most common to speech and music (Chapter 2). Here, we measure thresholds for spectral slope discrimination in adult listeners across a range of standard spectral slopes (–18 to +18 dB/octave), first, to determine how sensitive adult listeners are to changes in spectral slope, and second, to determine if adult listeners also show enhanced sensitivity in the range of speech and music, similar to that seen in infant listeners.

II. METHOD

A. Participants

Seventy undergraduate students (12 males, 58 females) whose mean age was 21 years (range 19 – 43 years) participated. All participants reported no hearing deficits. One participant was excluded from the analysis because of failure to reach criterion during the training block (see Procedure) and one other participant was excluded for near perfect performance (see Results).

B. Stimuli

Forty-nine complex tones with spectral slopes varying from –24 dB/octave to +24 dB/octave in 1 dB/octave steps were generated with a Symbolic Sound Corporation, Kyma Sound Design Workstation, version 4.5. Stimuli were generated with 16-bit precision at a sampling rate of 44.1 kHz and saved as digital sound files for stimulus presentation. All tones had a fundamental of 200 Hz, were 1000 ms in duration, including 20-ms linear rise and fall times, and each tone consisted of 5 harmonics in cosine phase. Sounds were presented at intensity levels ranging from 65 dB(A) to 77 dB(A) in 1 dB steps over a noise floor of 24 dB(A). This intensity variation of 12 dB was used to control for local intensity cues[2].

C. Apparatus

Testing was conducted in a sound-attenuating chamber (Industrial Acoustics Co.). The experiment was run using a Power Macintosh 7300/180 computer. The sound stimuli were passed through a Denon amplifier (PMA-480) to a single loudspeaker located inside the chamber. Loudspeaker presentation was used in order to compare the results with previous tests of infants. The participant sat in a chair facing the loudspeaker. Under the loudspeaker was a compartment containing lights which were used for feedback. The participant was 1.25 m away from the loudspeaker. A custom-built interface box connected a button box (used by the participant to signal a response) and response lights to the computer with a Strawberry Tree I/O card.

D. Procedure

A two-interval forced-choice method of constant stimuli procedure was used. Each trial consisted of two intervals, and each interval had two tones. The first tone of each interval had the standard spectral slope, and the second tone in the pair had either the standard tone again, or a tone with a different spectral slope. Whether the change was contained in Interval 1 or Interval 2 varied randomly across trials. Within a trial, the intensity of the first tones of each interval were the same, while the intensity of the second tone in each interval was randomly varied. Thus, three intensity levels were chosen randomly on each trial: the intensity of the two standards, the intensity of the second tone in the first interval, and the intensity of the second tone in the second interval (see Figure 1). Listeners were not instructed to listen for specific types of changes, but they were simply told that one of the intervals contained a change in tone quality. The participant's task was to press a button to indicate the interval in which the spectral slope (i.e., timbre) of the second tone differed (Interval 1 or Interval 2). The button box was connected to the computer which recorded the participant's response. There was a 1500 ms pause between intervals and each interval contained two tones separated by 1000 ms. Lights provided feedback on each trial. The next trial began 1000 ms after the listener made a response.

Each listener was randomly assigned to one of seven standard spectral slope conditions (–18, –12, –6, 0, +6, +12, +18 dB/octave). The slope value of the change tone varied in 1 dB/octave steps from 1 dB/octave above or below the standard to 6 dB/octave

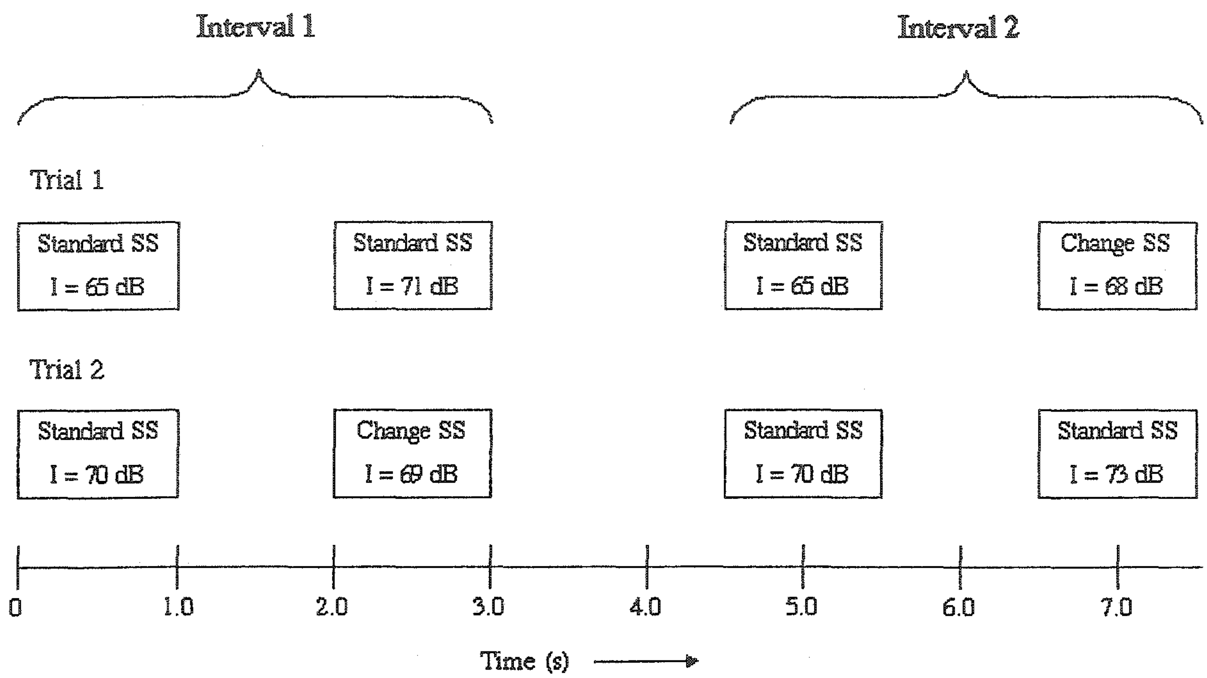


Figure 1. A schematic of two experimental trials. Each trial consists of two intervals (Interval 1 and Interval 2) separated by 1500 ms. Within a block, the standard tone always has the same spectral slope (one of –18, –12, –6, 0, +6, +12, +18 dB/octave) and the change tone always has the same spectral slope (1, 2, 3, 4, 5, 6 dB/octave above or below the standard spectral slope). However, the intensity varies from trial to trial. Each interval consists of two tones separated by 1000 ms. Within a trial, the first tones of the two intervals are identical in spectral slope (standard slope) and intensity, although the intensity varies randomly from trial to trial. The second tone is either the standard at a different intensity level, or the change at a different intensity level. The task of the participant is to choose the interval that contains the change, Interval 2 for the trial in the top panel, and Interval 1 for the trial in the lower panel.

above or below the standard. Trials were blocked by the 12 spectral slope changes with 24 trials in each block. Thus, each subject received a total of 288 trials across the entire testing session. We chose to test relatively few trials across a number of listeners in order to compare the results with those reported previously for infants. There were two orders of presentation: half of the participants received the trial blocks going from negative to positive changes (i.e., changes in order of presentation were: -6, -5, -4, -3, -2, -1, +1, +2, +3, +4, +5, +6 dB/octave), and the other half received the reverse order, going from positive to negative changes. Only two orders of presentation were used to minimize variation across participants, as well as between trial blocks within a participant.

At the beginning of the testing session, there was a set of training trials, included to give the listener practice in the task, as well as to demonstrate the sound quality change to listen for. During training trials, the standard tone had a –4 dB/octave slope while the change tone had a +4 dB/octave slope. In the training trials, there was no intensity variation, such that all four tones in the trial had the same intensity. In order to pass from the training phase to the experimental phase, the listener needed to give 5 correct answers in a row out of 20 trials. If the listener did not meet this criterion, the session was terminated (this was the case for only one participant, who was not included in the final sample).

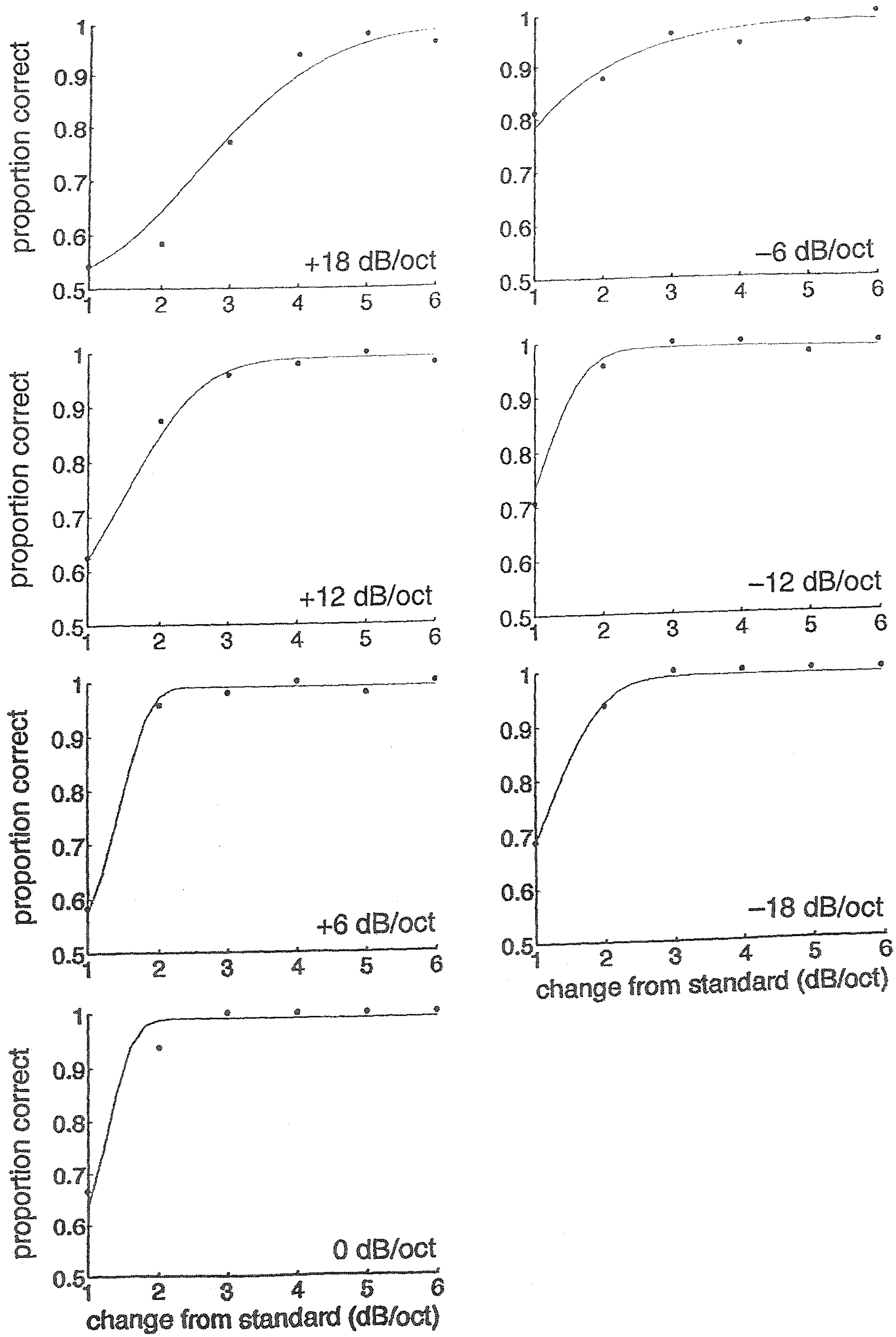
III. RESULTS

In order to determine thresholds, a Weibull function (e.g., Wichmann & Hill, 2001; Strausberger, 2001) varying from .51 to .99 was fitted to the proportion correct data for each individual participant, with separate functions derived for changes in the positive (i.e., +1, +2, +3, +4, +5, +6 dB/octave) and negative (i.e., -1, -2, -3, -4, -5, -6 dB/octave) directions. The function provided a good fit to the data for all participants with one exception. This participant, in the –6 dB/octave standard condition, had near perfect performance at all change sizes, and his data were excluded. Data and the corresponding Weibull fits are shown for a typical participant in each standard spectral slope condition in Figure 2. Because performance was quite high even with a change in slope of only 1 dB/octave, 85% thresholds were derived from the Weibull fits. Average thresholds are shown in Figure 3 for each standard spectral slope.

Separate thresholds were derived for changes above and below the standard spectral slope. An analysis of variance (ANOVA) was conducted with standard condition (–18, –12, –6, 0, +6, +12, +18 dB/octave) and order of presentation (trial blocks presented from positive to negative slope changes from the standard; trial blocks presented from negative to positive) as between-subjects factors, and direction of change (positive changes from standard; negative changes from standard) as a repeated measure. As predicted, there was a significant main effect of standard condition, $F(6, 56) = 4.63$, $p < 0.0007$. Figure 3 shows that average thresholds are lower for negative standard spectral slopes than for positive, with a minimum at –6 dB/octave. The only other significant effect was a main effect of order of presentation, $F(1, 56) = 10.41$, $p < 0.002$.

Figure 2 (next page). Psychometric functions for a typical subject in each standard condition. The solid line indicates the curve derived from the Weibull function, and points around the curve (solid dots) indicate the actual performance of the participant. Psychometric functions for standards with positive spectral slopes (0, +6, +12, +18 dB/octave) are shown in the left panels from top to bottom, while psychometric functions for standards with negative spectral slopes (–18, –12, –6 dB/octave) are shown in the right panels from top to bottom.

Participants who received negative to positive slope change trial blocks generally had higher performance than participants who received the reverse order of trial blocks, indicating that practice with downward changes appears to improve performance with upward changes more than the reverse.


proportion correct
+18 dB/oct
-6 dB/oct
+12 dB/oct
-12 dB/oct
+6 dB/oct
-18 dB/oct
0 dB/oct
change from standard (dB/oct)

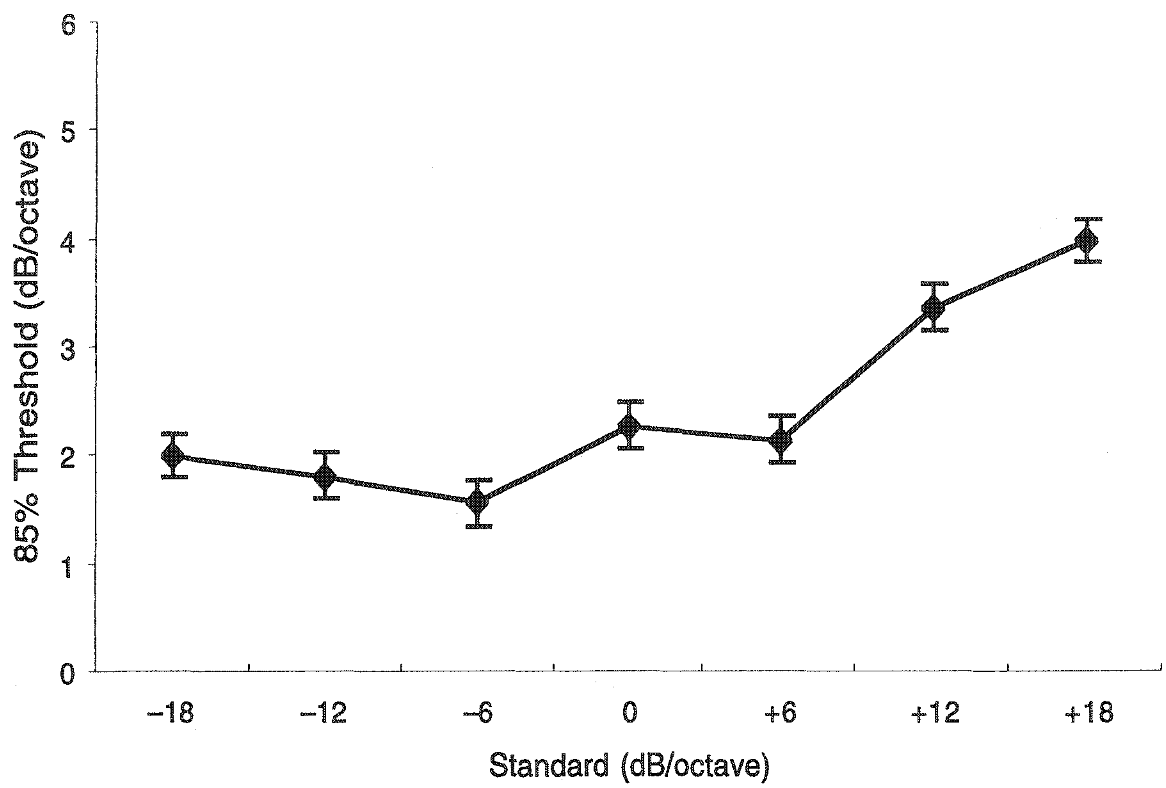


Figure 3. Average 85% correct thresholds across each standard condition, collapsed across direction of slope change. Standard error bars represent the standard error of the mean across observers.

IV. DISCUSSION

Spectral slope is a distinguishing feature of musical instruments, male versus female voices, and nasal versus non-nasal sounds. The results of this study show that adult listeners are able to use spectral slope as a dimension for timbre discrimination. For standards with negative spectral slopes, performance was high, with average 85% thresholds between 1 and 2 dB/octave. These thresholds are similar to those of the informal study with practiced listeners of Campbell (1994, note 3). Even for extremely positive spectral slopes (+12, +18 dB/octave), average 85% thresholds were 3 to 4 dB/octave. Like studies of profile analysis (e.g., Green & Kidd, 1982; Green, Kidd, & Picardi, 1983; Spiegel, Picardi, & Green, 1981), these results demonstrate that listeners are able to use global properties of the spectrum as a basis for sound discrimination.

The results of the present study also show that, as predicted, listeners are most sensitive to sounds with spectral slopes in the moderately negative range, with a minimum at –6 dB/octave. This result is consistent with infants' discrimination of spectral slope. Tsang & Trainor (2002) found that 8-month-old infants discriminated spectral slopes between –4 and –10 dB/octave, but were unable to discriminate between tones with positive spectral slopes differing by the same amount. These results suggest that the human auditory system may be tuned at an early age to be most sensitive to negative spectral slopes, and that this sensitivity is maintained into adulthood.

Two studies suggest that in the negative spectral slope region, spectral slope is processed independently from a number of other features contributing to the perceived timbre, such as rise time and fine spectral structure (Campbell, 1994; Li & Pastore, 1995). For the most part, speech and musical sounds have negative spectral slopes, and small differences in spectral slope within this region signal important information such as the identity, sex, and mood of a speaker. We found poorer spectral slope discrimination outside of the region of speech and music. It would be interesting to test whether spectral slope is also separable from other features contributing to timbre in this range, or whether this perceptual separability is restricted to the range encompassing environmentally important sounds.

Textual Footnotes

[1]ANSI, Psychoacoustics Terminology, s.3.20

[2]For discrimination of +18 dB/octave and +24 dB/octave spectral slopes (6 dB/octave difference), an intensity rove of 9.24 dB is need to cover local intensity cues that could potentially be used if listeners could hear out individual harmonics. This is the largest rove necessary across all possible standards and changes used. For further discussion of control of local intensity cues, see Tsang & Trainor (2002).

References

Beddor, P. S. (1993). "The perception of nasal vowels", in *Phonetics and Phonology, Vol. 5: Nasals, Nasalization and the Velum* edited by M. K. Huffman, R. A. Krakow, (Academic Press: San Diego, CA), pp. 171-196.

Bernstein, L. R., and Green, D. M. (1987). "Detection of simple and complex changes of spectral shape", Journal of the Acoustical Society of America, **82**, 1587-1592.

Campbell, S. L. (1994). "Uni- and multidimensional identification of rise time, spectral slope, and cutoff frequency", Journal of the Acoustical Society of America, **96**, 1380-1387.

Clarkson, M. G. (1996). "Infants' intensity discrimination: spectral profiles", Infant Behavior and Development, **19**, 181-190.

DeBruijn, A. (1978). "Timbre classification of complex tones", Acustica, **40**, 108-114.

Fletcher, N. H., and Rossing, T. D. (1981). *The Physics of Musical Instruments*. New York: Springer-Velag.

Green, D. M. (1983). "Successive versus simultaneous comparison in auditory intensity discrimination", Journal of the Acoustical Society of America, **73**, 639-643.

Green, D. M., and Kidd, G., Jr. (1982). "Discrimination of intensity increments to components of multitone complexes", Journal of the Acoustical Society of America, **71**, S37

Green, D. M., Kidd, G., Jr., and Picardi, M. C. (1983). "Successive versus simultaneous comparison in auditory intensity discrimination", Journal of the Acoustical Society of America, **73**, 639-643.

Grey, J. M. (1977) "Multidimensional perceptual scaling of musical timbres", Journal of the Acoustical Society of America, **61**, 1270-1277.

Grey, J. M., and Gordon, J. W. (1978). "Perceptual effects of spectral modifications on musical timbres", Journal of the Acoustical Society of America, **63**, 1493-1500.

Hall, D. E. (1980). *Musical Acoustics: An Introduction.* Belmont, CA: Wadsworth Publishing Co.

Hawkins, S. and Stevens, K. S. (1985). "Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels", Journal of the Acoustical Society of America, **77**, 1560-1575.

Iverson, P., and Krumhansl, C. L. (1993). "Isolating the dynamic attributes of musical timbre" Journal of the Acoustical Society of America, **94**, 2595-2603.

Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers", Journal of the Acoustical Society of America, **87**, 820-857.

Li, X., and Pastore, R. E. (1995). "Perceptual constancy of a global spectral property: spectral slope discrimination", Journal of the Acoustical Society of America, **98**, 1956-1968.

McAdams, S., and Cunible, J. C. (1992). "Perception of timbral analogies", Philosophical Transactions of the Royal Society, London, Series B, **336**, 383-389.

McAdams, S., Winsberg, S., Donnadieu, S., DeSoete, G., and Krimpoff, J. (1995). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes", Psychological Research, **58**, 177-192.

Miller, J. R., and Carterette, E. C. (1975) "Perceptual space for musical structures", Journal of the Acoustical Society of America, **58**, 711-720.

Olson, H. F. (1967). *Music, Physics and Engineering (2nd Ed.).* New York: Dover Publications.

Plomp, R., Pols, L. C. W., and Van de Geer, J. P. (1966) "Dimensional analysis of vowel spectra", Journal of the Acoustical Society of America, **41**, 707-712.

Pols, L. C. W., Van der Kamp, L. J. Th., and Plomp, R. (1969) "Perceptual and physical space of vowel sounds", Journal of the Acoustical Society of America, **46**, 458-467.

Scherer, K. R. (1989). "Vocal correlates of emotional arousal and affective disturbance", In *Handbook of Social Psychology* edited by H. Wagner and A. Manstead (John Wiley & Sons, Ltd).

Spiegel, M. F., Picardi, M. C., and Green, D. M. (1981). "Signal and masker uncertainty in intensity discrimination", Journal of the Acoustical Society of America, **70**, 1015-1019.

Strausberger, H. (2001). "Converting between measures of slope of the psychometric function", Perception and Psychophysics, **63**, 1348-1355.

Sundberg, J. (1991). *The Science of Musical Sounds*, San Diego, CA: Academic Press

Tsang, C. D., and Trainor, L. J. (2002). "Spectral slope discrimination in infancy: sensitivity to socially important timbres", Infant Behavior and Development, **25**, 183-194.

Versfeld, N. J., and Houtsma, A. J. M. (1991) "Perception of spectral changes in multi-tone complexes", The Quarterly Journal of Experimental Psychology, **43A**, 459-470.

Wichmann, F. A., & Hill, N. J. (2001) "The psychometric function: I. Fitting, sampling, and goodness of fit, Perception and Psychophysics, **63**, 1293-1313.

CHAPTER 5

GENERAL DISCUSSION

The studies presented in this thesis reveal a number of important findings. First, 8-month-old infants are sensitive to changes in spectral slope in a limited range around –10 dB/octave and –4 dB/octave (Chapter 2). Second, 8-month-olds show an attentional bias for sounds with spectral slopes in this range (Chapter 3). Third, sensitivity to spectral slope has a developmental progression, in that it appears to emerge around 5 months of age (Chapter 3). Fourth, spectral slope discrimination is much better in adults than in infants, but adults also show optimal sensitivity for spectral slopes in the moderately negative range, similar to infant listeners (Chapter 4).

These results support the notion that global properties of the spectrum can be used by listeners as a basis for timbre discrimination. This is consistent with studies of profile analysis, in which listeners have been shown to be more sensitive to relative changes in spectral shape than to absolute intensity changes in individual components (Green, 1983; Green & Kidd, 1982; Green, et al., 1983; Speigel, et al., 1981; Versfeld & Houtsma, 1991). The studies presented here extend the findings of the profile analysis literature. Adult listeners are able to detect very small differences in spectral slope, with best discrimination in the negative spectral slope region compared to the positive spectral slope region. Furthermore, 8-month-old infants are also maximally sensitive to spectral slopes in this region.

Spectral slope has been found to be a distinguishing feature of male and female voices (e.g., Klatt & Klatt, 1990), as well as vocal quality identification (e.g. Beddor, 1993), and the recognition of different vocal expressions of emotion (Scherer, 1989). The studies presented here also support the idea that spectral slope may be an important perceptual cue in speech and music, as moderately negative spectral slopes are most commonly found in speech and music. This is also consistent with the view proposed by Li & Pastore (1995), which related spectral slope to source characteristics of voices. Source characteristics, like spectral slope, remain constant across changing filter characteristics (such as number of frequency components, location of resonances). Thus, spectral slope may be a useful cue to recognizing the same voice across varying speech sounds, and in a broader sense, may be useful in auditory object recognition.

The development of spectral slope discrimination may be influenced by a number of factors. The data presented here show that sensitivity to spectral slope develops around 5 months of age, because females in this age range are able to discriminate tones with positive from tones with negative spectral slopes, but males do not show evidence of this discrimination (Chapter 3). By 8-months, both sexes show enhanced discrimination abilities (Chapter 2) and also show attentional biases for moderately negative spectral slopes (Chapter 3). Thus, there is support for the notion that infant attentional preferences may result from discrimination abilities and experience, as discrimination of spectral slope appears to arise before preference for negative spectral slopes. The attentional preference and enhanced sensitivity for moderately negative spectral slopes may provide infants with a means to attend more to sounds with negative spectral slopes, such as speech and music, over other stimuli. Interestingly, relevant language-specific abilities show the most rapid development between 5 to 8 months, both in terms of the sounds infants babble and infants' perception of language-specific speech sound categories (see deBoysson-Bardies, et al., 1992). Thus, the developmental changes in spectral slope perception also parallel the development of linguistic abilities in infancy.

Limitations and Future Considerations

This thesis clearly demonstrates that infant and adult listeners are able to use the global dimension of spectral slope as a means for timbre discrimination. However, this thesis used artificially generated complex tones as stimuli in all the studies presented. Although spectral slope is related to a number of different important auditory processes, from voice and object recognition to identification of vocal expressions of emotion, a number of other cues are also present in real world stimuli that may be used in conjunction with, or supercede, the use of spectral slope differences as a means of timbre discrimination. Timbre is highly complex. Temporal characteristics of sound are also a distinguishing feature of many musical instrument timbres (Grey, 1977; McAdams et al., 1995; Cutting & Rosner, 1974), and dynamic spectral-temporal aspects of sounds have been shown to play a critical role in the identification and recognition of consonants in speech for adult listeners (Liberman, et al., 1961; Cutting & Rosner, 1974) and infant listeners (Jusczyk, et al., 1977). In the studies presented in this thesis, the stimuli were designed such that discriminations based on spectral slope differences would result in best performance. Thus, although it is clear from the studies presented in this thesis that adult and infant listeners alike are sensitive to global spectral properties, it remains for future research to examine the extent to which spectral slope is actually used by listeners in real world auditory discriminations, such as with voices and musical instrument tones.

The extent to which sex differences are relevant to the discrimination of spectral slope may also be a question for future research. Female 5-month-old infants discriminate between highly different spectral slopes (positive and negative) whereas male 5-month-olds do not (Chapter 3). This finding, in conjunction with evidence that female infants show superior performance in language related processes (Darley & Winitz, 1961; Rome-Flanders & Cronk, 1995), supports the notion that discrimination of spectral slope occurs before an attentional bias to environmentally common spectral slopes. However, it may also be the case that female infants show earlier development of motor or attentional abilities, which may enhance performance on an operantly conditioned head-turn task. Therefore, future research should attempt to replicate the results of Experiment 3 in Chapter 3 using measures that do not require infants to make head-turns, such as observer-based or electrophysiological methods.

Furthermore, the results of the infant studies (Chapters 2 and 3) do not explicitly examine the role of innate and experiential factors in the development of spectral slope sensitivity. It is clear that the discrimination of spectral slope arises before attentional preferences to negative spectral slope (Chapter 3). However, it is not known whether innate factors are responsible for the enhanced discrimination abilities, or if exposure to real-world sounds (most of which have negative spectral slopes) leads to enhanced sensitivity, and ultimately an attentional bias. As noted in Chapter 2, one way to examine this issue would be to provide infants with increased exposure to sounds with rare, positive spectral slopes and determine if this exposure would increase the discriminability of positive spectral slopes.

Conclusions

The results presented here support studies of timbre perception showing that global characteristics of the spectral envelope can play a large role in timbre perception. Adult listeners and 8-month-old infants show enhanced discrimination abilities for sounds with spectral slopes in the range of speech and music, and furthermore, there appears to be age-related changes in the discrimination of spectral slope. Spectral slope is an important cue for sound recognition, identification and discrimination. The development of spectral slope sensitivity is one potentially important aspect of auditory development, and may reflect both innate and experience-dependent factors.

CHAPTER 1 and 5 REFERENCES

Beddor, P. S. (1993). The perception of nasal vowels. In Phonetics and Phonology, vol. 5: Nasals, Nasalization and the Velum edited by M. K. Huffman, R. A. Krakow, San Diego CA: Academic Press, p. 171-196.

Bernstein, L. R., & Green, D. M. (1987). Detection of simple and complex changes of spectral shape. Journal of the Acoustical Society of America, 82, 1587-1592.

VonBismark, G. (1974). Timbre of steady sounds: A factorial investigation of its verbal attributes. Acustica, 30, 146-159.

Bloom, K., Moore-Schoenmakers, & Mastaka, N. (1999). Nasality of infant vocalizations determines gender bias in adult favorability ratings. Journal of Nonverbal Behavior, 23, 219-235.

Campbell, S. L. (1994). Uni- and multidimensional identification of rise time, spectral slope, and cutoff frequency. Journal of the Acoustical Society of America, 96, 1380-1387.

Clarkson, M. G. (1996). Infants' intensity discrimination: spectral profiles. Infant Behavior and Development, 19, 181-190.

Clarkson, M. G., Clifton, R. K., & Perris, E. E. (1988). Infant timbre perception: Discrimination of spectral envelopes. Perception and Psychophysics, 16, 564-570.

Cutting, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music. Perception and Psychophysics, 16, 564-570.

Darley, F. L., & Winitz, H. (1961). Age of first word: Review of research. Journal of Speech and Hearing Disorders, 26, 272-290.

DeBruijn, A. (1978). Timbre classification of complex tones, Acustica, 40, 108-114.

DeBoysson-Bardies, D., Viham, M. M., Roug-Hillichius, L., Durand, C., Landberg, I., & Arao, F. (1992). Material evidence of infant selection from the target language: a cross linguistic phonetic study. In C. A. Ferguson, L. Menn, C. Stoel-Gammon (Eds.) *Phonetic Development: Models, Research, Implications*, Timonium, Maryland: York Press, Inc., pp. 369-391.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. Science, 208, 1174-1176.

Fletcher, N. H., & Rossing, T. D. (1981). The Physics of Musical Instruments. New York: Springer-Velag.

Green, D. M. (1992). The number of components in profile analysis tasks. Journal of the Acoustical Society of America, 91, 1616-1623.

Green, D. M., & Kidd, G., Jr. (1982). Discrimination of intensity increments to components of multitone complexes. Journal of the Acoustical Society of America, 73, 639-643.

Green, D. M., Kidd, G., Jr, & Picardi, M. C. (1983). Successive versus simultaneous

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America. 61, 1270-1277.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. Journal of the Acoustical Society of America, 63, 1493-1500.

Hall, D. E. (1980). Musical Acoustics: An Introduction. Belmont, CA: Wadsworth Publishing Co.

Hawkins, S., & Stevens, K. S. (1985). Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels. Journal of the Acoustical Society of America, 77, 1560-1575.

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. Journal of the Acoustical Society of America, 94, 2595-2603.

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. F., & Smith, L. B. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. Perception and Psychophysics, 21, 50-54.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America, 87, 820-857.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America, 66, 1668-1679.

Li, X., & Pastore, R. E. (1995). Perceptual constancy of a global spectral property: spectral slope discrimination. Journal of the Acoustical Society of America. 98, 1956-1968.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset time of the components of certain speech and nonspeech patterns. Journal of Experimental Psychology, 61, 379-388.

Marean, G. C., Werner, L. A., & Kuhl, P. K. (1992). Vowel categorization by very young infants. Developmental Psychology, 28, 396-405.

McAdams, S., & Cunible, J. C. (1992). Perception of timbral analogies, Philosophical Transactions of the Royal Society, London, Series B, 336, 383-389.

McAdams, S., Winsberg, S., Donnadieu, S., DeSoete, G., & Krimpoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. Psychological Research, 58, 177-192.

Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. Infant Behavior and Development, 6, 313-330.

Miller, J. R., and Carterette, E. C. (1975). Perceptual space for musical structures. Journal of the Acoustical Society of America, 58, 711-720.

Murray, I. R., & Arnott, J. L. (1993). Toward simulation of emotion in synthetic speech, A review of the literature on human vocal emotion. Journal of the Acoustical Society of America, 93, 1097-1108.

Olsho, L.W., Schoon, C., Sakai, R., Turpin, R., & Sperduto, V. (1982a). Preliminary data on frequency discrimination in infancy. Journal of the Acoustical Society of America, 71, 509-511.

Olsho, L.W., Schoon, C., Sakai, R., Turpin, R., & Sperduto, V. (1982b). Auditory frequency discrimination in infancy. Developmental Psychology, 18, 721-726.

Olson, H. F. (1967). Music, Physics, and Engineering (2nd Ed.). New York: Dover Publications.

Plomp, R., Pols, L. C. W., & Van de Geer, J. P. (1966). Dimensional analysis of vowel spectra. Journal of the Acoustical Society of America, 41, 707-712.

Pols, L. C. W., Van der Kamp, L. J. Th., & Plomp, R. (1969). Perceptual and physical space of vowel sounds. Journal of the Acoustical Society of America, 46, 458-467.

Rome-Flanders, T., & Cronk, C. (1995). A longitudinal study of infant vocalizations during mother-infant games. Journal of Child Language, 22, 259-274.

Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner & A. Manstead (Eds.), Handbook of Social Psychophysiology. John Wiley & Sons Ltd.

Sinnott, J.M., & Aslin, R.N. (1985). Frequency and intensity discrimination in human infants and adults. Journal of the Acoustical Society of America. 78. 1986-1992.

Speigel, M. F., Picard, M. C., & Green, D. M. (1981). Signal and masker uncertainty in intensity discrimination. Journal of the Acoustical Society of America, 70, 1015, 1019.

Sundberg, J. (1991). The Science of Musical Sounds, San Diego, CA: Academic Press.

Trehub, S. E., Endman, M. W., & Thorpe, L. A. (1990). Infants' perception of timbre: Classification of complex tones by spectral structure. Journal of Experimental Child Psychology, 49, 300-313.

Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of instrumental timbre. Scandanavian Journal of Psychology, 13, 228-240.

Werner, L., & Marean, G. C. (1996). Human Auditory Development. Madison, WI: Brown & Benchmark Publishers.

Versfeld, N. J., & Houtsma, A. J. M. (1991). Perception of spectral changes in multitone complexes. The Quarterly Journal of Experimental Psychology, 43A, 459-470.