UNIVERSITY OF ALBERTA

PERCEPTUAL AND ACOUSTIC ANALYSIS OF WORD-INITIAL VOICING CONTRASTS ACROSS SPEAKER AGE

BY

KELLY WYNNE LUCKY

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science in Speech-Language Pathology.

DEPARTMENT OF SPEECH PATHOLOGY AND AUDIOLOGY

Edmonton, Alberta

Spring, 1993

ISBN 0-315-82212-0

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Kelly Wynne Lucky

TITLE OF THESIS: Perceptual and acoustic analysis of word-initial voicing contrasts across speaker age.

DEGREE: Master of Science in Speech-Language Pathology

YEAR THIS DEGREE GRANTED: 1993

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Kelly Lucky

9208-71 St., Edmonton, AB, T6B 1Y2

DATE: 11 January '93

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **PERCEPTUAL AND ACOUSTIC ANALYSIS OF WORD-INITIAL VOICING CONTRASTS ACROSS SPEAKER AGE** submitted by Kelly Wynne Lucky in partial fulfillment of the requirements for the degree of Master of Science, Department of Speech-Language Pathology and Audiology.

Anne Putnam Rochet
Dr. Anne Putnam Rochet

Megan M. Hodge
Dr. Megan Mary Hodge

Sharon Warren
Dr. Sharon Warren

DATE: 18 December '92

Abstract

The purpose of this study was two fold: 1) to compare the values of four acoustic characteristics of word-initial alveolar plosives produced by normal speakers of Canadian English, and 2) to examine the individual and combined effects of these acoustic characteristics on listeners' perceptions of the voicing contrast. Speakers at four age levels participated in the study: 2 years 6 months to 3 years; 4 years 6 months to 5 years; 10 years to 11 years; and adults. Speakers were audiotape-recorded as they produced word-initial /t/ and /d/ in minimal-pair monosyllabic words. Five English-speaking adult listeners judged the word-initial plosives of the speakers' recorded productions as voiced, voiceless, or ambiguous with respect to voicing. The speakers' taped utterances were digitized using microcomputer-based speech waveform analysis procedures and measured for four acoustic characteristics: Voice onset time, first formant frequency at the onset of the vowel, fundamental frequency at the onset of the vowel, and plosive burst amplitude. The data were analyzed using analyses of variance and a simple discriminant function analysis.

Acoustic data analysis indicated that for all speakers VOT was longer, F1 onset frequency was higher, and burst amplitude was greater for /t/ than /d/. Only measures of F0 did not significantly differentiate voiced and voiceless plosives. Spectral measures were higher for the younger speakers even when a spectral normalization procedure was applied. Variability within all four acoustic measures was generally highest for the youngest subjects suggesting that speakers became more consistent with age in their productions of the acoustic parameters.

Perceptual data analysis revealed that although all speakers produced perceptually-distinct voicing contrasts, listeners' abilities to accurately perceive this contrast improved as speaker age increased. The results of the discriminant function analysis supported the argument for VOT as a primary cue for the perception of voicing, at least for perceptually-validated word-initial alveolar plosives. Spectral information at the onset of the vowel, and burst amplitude to a lesser extent, provided secondary information to aid the listeners' perceptions of these plosives. Future studies are required to determine whether these relationships are influenced by speaker age, perceptual ambiguity, and phonetic context.

Acknowledgements

A large number of individuals have contributed to this study. Firstly, I thank all of the subjects that volunteered their time and energies to participate. As many of the subjects were of a young age, their participation necessarily required the involvement of many parents and daycares to whom I am also grateful.

I appreciate the assistance and support that I received for the [illegible] of this project from my employer, the Glenrose Rehabilitation H[illegible] I thank Elaine Heaton, Director of the Department of Communication Disorders, for the flexibility that I was permitted in scheduling my time, and for unlimited access to the microcomputer speech analysis system. I am also thankful to Dr. Kerrie Pain for her assistance in developing and interpreting the statistical analyses.

I am also indebted to the various people and departments at the University of Alberta. The Department of Linguistics allowed me use of their facility, hardware and software required for the perceptual task. In this regard, I acknowledge the great assistance given to me by Dr. Bernard Rochet in converting my data to the format necessary to run the perceptual analysis. Additionally, the Centre for Research in Applied Measurement and Evaluation, and Michelle Samuel particularly, were indispensable in the help that they provided to input and analyze the data.

Finally, I am deeply grateful to my Committee members, Dr. Anne Rochet, Dr. Megan Hodge, and Dr. Sharon Warren, for their invaluable insights, and limitless patience and understanding throughout this process.

TABLE OF CONTENTS

INTRODUCTION 1

LITERATURE REVIEW 2

VOICE ONSET TIME 3

DEFINITION 3

ADULT PRODUCTION DATA 3

ADULT PERCEPTUAL DATA 4

DEVELOPMENTAL CHANGES IN VOT 5

CONTEXTUAL FACTORS INFLUENCING VOT 6

FIRST FORMANT TRANSITION 8

DEFINITION 8

EFFECT OF F1 CHANGES ON VOICING CONTRAST . . . 10

SUMMARY 15

FUNDAMENTAL FREQUENCY 16

DEFINITION 16

DEVELOPMENTAL CHANGES IN F0 16

EFFECT OF F0 CHANGES ON VOICING CONTRAST . . . 16

SUMMARY 20

PLOSIVE BURST CHARACTERISTICS 20

DEFINITION 20

EFFECTS OF PLOSIVE BURST ON VOICING 22

SUMMARY OF ACOUSTIC CUES 23

STUDIES OF VOICING USING PERCEPTUAL VALIDATION TASKS 25

LIMITATIONS OF ACOUSTIC ANALYSIS 29

SPECTRAL NORMALIZATION 32

PURPOSE 35

METHOD 42

DESIGN 42

SUBJECTS 44

PROCEDURES 48

PRODUCTION TASK 48

ACOUSTICAL ANALYSES 51

VOICE ONSET TIME 53

FIRST FORMANT ONSET FREQUENCY 53

FUNDAMENTAL FREQUENCY 57

PLOSIVE BURST AMPLITUDE 57

PERCEPTUAL ANALYSIS 60

RELIABILITY 63

INTRA-LISTENER RELIABILITY 63

ACOUSTIC MEASURES RELIABILITY 63

DATA ANALYSIS 65

PERCEPTUAL DATA 65

INTRA-LISTENER RELIABILITY 66

INTER-LISTENER COMPARISONS 67

ACOUSTICAL DATA 69

RESULTS 75

PERCEPTUAL DATA 75

INTRA-LISTENER RELIABILITY 75

INTER-LISTENER COMPARISONS 82

ACOUSTIC DATA 96

ACOUSTIC MEASURES RELIABILITY 96

COMPARISONS AMONG ACOUSTIC VARIABLES 98

DISCRIMINANT FUNCTION ANALYSIS 112

DISCUSSION.....116

FINDINGS AND IMPLICATIONS.....118

ACOUSTIC ANALYSIS.....118

SPECTRAL NORMALIZATION.....118

VOICE ONSET TIME.....120

FIRST FORMANT.....123

FUNDAMENTAL FREQUENCY.....127

PLOSIVE BURST AMPLITUDE.....130

PERCEPTUAL ANALYSIS.....131

INTRA LISTENER RELIABILITY.....131

INTER-LISTENER COMPARISONS.....132

CONTRIBUTION OF THE ACOUSTIC CUES TO LISTENERS' PERCEPTIONS.....138

EXPERIMENTAL HYPOTHESES REVISITED.....141

THREATS TO VALIDITY OF THE RESULTS.....145

INTERNAL VALIDITY.....145

EXTERNAL VALIDITY.....153

DIRECTIONS FOR FUTURE STUDY.....155

CONCLUSIONS.....157

REFERENCES.....159

APPENDIX A: CALCULATION OF SAMPLE SIZE.....168

APPENDIX B: SUMMARY TABLE OF POTENTIAL SUBJECTS WHO DID NOT PASS THE SCREENING PROCEDURES.....169

APPENDIX C: WORD LIST FOR PRODUCTION TASK.....170

APPENDIX D: FORMULA FOR BARK TRANSFORMATION.....171

APPENDIX E: GROUP MEANS AND STANDARD DEVIATIONS FOR VOT (IN MS) ACCORDING TO SPEAKER AGE, VOICING CONDITION, AND GENDER........................172

APPENDIX F: GROUP MEANS AND STANDARD DEVIATIONS FOR FIRST FORMANT (F1) AT ONSET OF THE VOWEL /ɑ/ ACCORDING TO SPEAKER AGE, VOICING CONDITION, AND GENDER.....................................173

APPENDIX G: GROUP MEANS AND STANDARD DEVIATIONS FOR FUNDAMENTAL FREQUENCY (F0) AT ONSET OF THE VOWEL /ɑ/ ACCORDING TO SPEAKER AGE, VOICING CONDITION, AND GENDER........................174

APPENDIX H: GROUP MEANS AND STANDARD DEVIATIONS FOR RATIOS OF PLOSIVE BURST AMPLITUDE RELATIVE TO VOWEL AMPLITUDE (IN VOLTS) ACCORDING TO SPEAKER AGE, VOICING CONDITION, AND GENDER..............175

APPENDIX I: CALCULATION OF EFFECT SIZES..................176

LIST OF TABLES

Table 1 page 76

Intra-rater reliability scores (%) for complete agreements for each listener across speaker age and by voicing condition.

Table 2 page 77

Intra-rater reliability scores (%) for within-category agreements for each listener across speaker age and by voicing condition.

Table 3 page 80

Mean scores and standard deviations of complete agreements among listeners across speaker age and between voicing condition.

Table 4 page 81

Mean scores and standard deviations of within-category agreements among listeners across speaker age and between voicing condition.

Table 5 page 83

Mean scores and standard deviations of listeners' correct judgments across speaker age and between voicing conditions.

Table 6 page 84

Mean scores and standard deviations of listeners' incorrect judgments across speaker age and between voicing conditions.

Table 7 page 85

Mean scores and standard deviations of listeners' ambiguous judgments across speaker age and between voicing conditions.

Table 8 page 99

Group means and standard deviations for voice onset time (in ms) across speaker age and between voicing conditions.

Table 9 page 100

Group means and standard deviations for first formant (F1) at onset of the postconsonantal vowel across speaker age and between voicing conditions.

Table 10 page 101

Group means and standard deviations for fundamental frequency (F0) at onset of the postconsonantal vowel across speaker age and between voicing conditions.

Table 11 page 102
Group means and standard deviations for ratios of plosive burst amplitude relative to vowel amplitude (in Volts) across speaker age and between voicing condition.

Table 12 page 124
Mean F1 values (in Hz) and standard deviations for /ɑ/ produced by speakers of comparable ages reported in various studies.

LIST OF FIGURES

Figure 1 page 54
Example of VOT measurement for /d/ on a waveform of /dɑt/ spoken by an adult male.

Figure 2 page 56
Example of F1 measurement from FFT and LPC spectra of a waveform of /ɑ/ in /dɑt/ spoken by an adult male.

Figure 3 page 58
Example of F0 measurement for one pitch period of a waveform of /ɑ/ in /dɑt/ spoken by an adult male.

Figure 4 page 59
Example of amplitude measurement of the plosive burst for /d/ in /dɑt/ spoken by an adult male.

Figure 5 page 88
Mean scores of correct judgments of /t/ and /d/ for five listeners across speaker age and between voicing conditions.

Figure 6 page 89
Mean scores of incorrect judgments of /t/ and /d/ for five listeners across speaker age and between voicing conditions.

Figure 7 page 90
Mean scores of ambiguous judgments of /t/ and /d/ for five listeners across speaker age and between voicing conditions.

Figure 8 page 92
Mean scores of correct judgments of /t/ and /d/ for four listeners across speaker age and between voicing conditions.

Figure 9 page 93
Mean scores of incorrect judgments of /t/ and /d/ for four listeners across speaker age and between voicing conditions.

Figure 10 page 94
Mean scores of ambiguous judgments of /t/ and /d/ for four listeners across speaker age and between voicing conditions.

Figure 11 page 107
Mean VOT values (in ms) of perceptually-validated tokens of /dɑt/ and /tɑt/ across speaker age and between voicing conditions.

Figure 12 page 108
Mean F1 values (in log) of perceptually-validated tokens of /dɑt/ and /tɑt/ across speaker age and between voicing conditions.

Figure 13 page 109
Mean F0 values (in log) of perceptually-validated tokens of /dɑt/ and /tɑt/ across speaker age and between voicing conditions.

Figure 14 page 110
Mean plosive burst amplitude (expressed as a ratio of plosive burst amplitude to vowel amplitude, in Volts) of perceptually-validated tokens of /dɑt/ and /tɑt/ across speaker age and between voicing conditions.

Introduction

The relationship between listeners' perceptions of the voicing contrast and the contribution of acoustic cues influencing perception of this contrast, is complex. There are several acoustic cues, including voice onset time, first formant frequency, fundamental frequency, and burst amplitude, that influence how a listener perceives a plosive with respect to the feature of voicing. These cues may complement each other, providing the listener with a redundant cuing system contributing to accurate perception. Conversely, these cues may provide contradictory information causing the listener to be uncertain in his judgment. Additionally, the relative importance of the acoustic cues may vary depending on the contextual information available to the listener.

The normal adult speaker has had years of practice at manipulating his neuromuscular system in a consistent manner, thereby providing consistent acoustical cues to the listener. What about the child speaker? Immature control of his speech system may result in variable production of one or more of these acoustic cues. Limited information is available regarding children's production of these cues and the way in which these cues influence the listener's perception of the voicing contrast in children's speech.

As yet, there exists neither a data base for average values of the acoustical cues thought to contribute to the perception of voicing in the speech of normal children across various age levels, nor reliable information about the relative contribution of these acoustical parameters to listeners' perceptions of the voicing contrast. This information would be particularly helpful as a basis for comparisons among speakers of various ages and speakers with various phonological disorders. This study provides preliminary data on four acoustic characteristics of normally-developing English speakers' productions of alveolar plosives produced in single words, and explores the relationship between these acoustic characteristics and listeners' perceptions of the voicing contrast in the speakers' productions of these plosives.

Literature Review

As mentioned previously, several acoustic cues are thought to influence listeners' perceptions of the voicing contrast in word-initial plosives: namely, voice onset time, first formant frequency at the onset of the postconsonantal vowel, fundamental frequency at the onset of the postconsonantal vowel, and plosive burst amplitude. The following discussion includes definitions of these acoustic characteristics and summarizes the developmental changes that occur during children's acquisition of speech motor control. Studies investigating the nature of these acoustic cues as a function of word-initial plosive voicing, and the impact of experimental manipulation of these cues on listeners' perceptions of voicing, are also summarized. Finally, limitations of acoustic measurement for certain features of children's speech and possible remedies to alleviate these difficulties are discussed.

Voice Onset Time

Definition.

Voice onset time (VOT) is an acoustical measure of the voicing distinction between voiced and voiceless speech sounds. VOT for plosives has been defined as the time interval between the release burst of the plosive and the onset of glottal pulsing for the following vowel or semivowel (Lisker & Abramson, 1964). The adult production of voiceless stops in English is characterized by a relatively long lag between the plosive burst and the onset of phonation for the following vowel. Conversely, English voiced stops are produced with a shorter lag (phonation begins simultaneously with the release burst or within a few milliseconds of it) or are prevoiced (phonation begins prior to the plosive burst) (Lisker & Abramson, 1964, 1967).

Adult production data.

Although variability in VOT for a given speech sound occurs within and between subjects, there is a well-established range of normal VOT for voiced and voiceless plosives. English voiced plosives are characterized by a VOT range from 0 to +40 ms; voice onset times for English voiceless plosives range from approximately +50 to +110 ms (Zlatin & Koenigsknecht, 1976). There is minimal overlap of the VOT ranges for voiced and voiceless plosives.

Adult perceptual data.

The production data for VOT correspond closely to the perceptual boundaries established between voiced and voiceless English plosives, identified between +20 and +40 ms (Abramson & Lisker, 1968; Zlatin & Koenigsknecht, 1975). The particular boundary depends on the plosive being identified, with longer VOT values associated with progressively posterior places of articulation.

If VOT is prolonged by more than 25 to 40 ms (dependent on the plosive) relative to the plosive release, the plosive burst and the onset of voicing are perceived as two distinct events and a voiceless plosive will likely be perceived. Conversely, if VOT is less than approximately 20 ms, the plosive burst and the voicing onset will be perceived as simultaneous events leading to the perception of a voiced plosive (Klatt, 1975). While this perceptual distinction holds true for most adult productions of plosives, other segmental and suprasegmental factors may override or complicate the process of accurate identification. (Factors affecting perception of the voicing distinction will be discussed in greater detail in a later section.)

Developmental changes in VOT.

The acquisition of the voicing contrast by children, as measured by VOT, has been well documented (Barton & Macken, 1980; Eguchi & Hirsh, 1969; Gilbert, 1977; Kewley-Port & Preston, 1974; Macken & Barton, 1979; Zlatin & Koenigsknecht, 1976). Children progress through three distinct stages:

(a) All stops are produced within the short lag range of the VOT continuum and are indistinguishable from one another.

(b) VOT ranges begin to assume a bimodal distribution, but VOT values still fall within the range characteristic of adult voiced stops. Data reported by Macken and Barton (1979) indicate that the contrast produced by children entering this stage may not initially be perceptible to an adult listener.

(c) The ranges of VOT for voiced vs. voiceless consonants assume adult categorical values. Macken and Barton (1979) suggested that this final stage in VOT acquisition may be further subdivided. At less mature stages of development, VOT values of both voiceless and voiced plosives may be significantly longer than comparable adult productions of these phonemes. As the child refines his control of the laryngeal gesture necessary to initiate or delay voicing, VOT values are shortened to approximate those of adult productions.

Additionally, there is significantly greater variability among VOT values for plosives produced by children in comparison to those produced by adults, particularly for voiceless plosives. Variability decreases with increasing age. By approximately 8 years, children exhibit VOT distributions that correspond to those of adult productions (Eguchi & Hirsh, 1969; Kent, 1976).

Contextual factors influencing VOT.

Voice onset time values vary within and across speakers due to a variety of contextual factors (Kent, 1976). Influential suprasegmental factors include speech rate, utterance length (Lisker & Abramson, 1967), and word stress (Klatt, 1975; Lisker & Abramson, 1967). For example, word-initial plosives in connected utterances are produced with minimal aspiration and shorter VOTs than the initial plosive of a single-word utterance. Consequently, VOT values of word-initial plosives in sentence contexts may be similar to those for intervocalic plosives and quite different from word-initial plosives produced in isolated words.

The phonetic context may also alter the VOT values of a given plosive. For example, vowel context influences the VOT values of prevocalic plosives such that VOT values are lengthened for both voiced and voiceless plosives when they are followed by high [illegible]s /i,u/ compared to low vowels /ɑ,ɔ/ (Klatt, 1975; Weismer, 1981).

The point of this discussion has been to highlight the complex interaction of production factors that influence VOT and affect the listener's perception of voiced and voiceless plosives. VOT is one of many variables, contextual and acoustic, that influences the perception of the voicing contrast. In order to effectively study the influence of acoustic variables, such as VOT, without the potentially confounding effects of context, the experimental task must be carefully controlled.

As the writer alluded to previously, several other acoustic cues contribute to the perception of the voicing contrast in word-initial plosives. They may vary systematically with changes in VOT or, in some cases, over-ride the perceptual information provided by VOT. Taken together, they may provide the listener with increased redundancy in the signal allowing for an accurate perception of the plosive (Borden & Harris, 1980). Conversely, these cues may be contradictory resulting in a conflicting message to the listener. For example, Summerfield and Haggard (1977) provided data suggesting that changes in F1 onset frequencies influence the perception of the voicing contrast and actually may over-ride the perceptual importance of the VOT information. According to Klatt (1975), the acoustic variables that cue a voicing distinction for word-initial plosives in association with VOT include (a) the frequency of the first formant at the onset of voicing and the associated spectral changes during the F1 transition, (b) fundamental frequency changes for the vowel onset following the plosive, and (c) the amplitude and duration of the plosive burst.

First Formant Transitions

Definition.

Baken (1987) defined a formant frequency as "a single frequency at which vocal tract transmission is more efficient than at nearby frequencies" (p.357)
The voice signal has significant energy only at discrete harmonics of the fundamental frequency. As the glottal source and vocal-tract filter are essentially independent, formant frequencies do not necessarily coincide directly with specific harmonic frequencies. The formant frequencies for a particular speaker must be estimated from the amplitudes of the harmonics. The harmonics close to the resonating frequencies of the vocal tract will be transmitted with optimal amplitude conservation, while harmonics that do not will be transmitted less efficiently.

Any change in vocal tract shape or size will result in a change of its resonating frequencies. Obviously, developmental changes in the size and shape of the vocal tract will effect changes in the formant frequencies. Generally, the smaller vocal tracts of adult females and children will be characterized by higher formant frequencies, whereas the larger vocal tracts of adult males will be characterized by lower formant values. Age also influences the stability of formant frequency values. Variability within any one formant frequency decreases with increasing subject age. By age 11 years, a speaker has usually achieved adult like stability (Kent, 1976).

The vocal tract is a variable resonator; therefore, within speaker changes in the resonating frequencies will result from changes in the shape of the vocal tract due to movement of the articulators. The point of maximum constriction of the vocal tract and the length of the tract from the glottis to the point of maximum constriction have the greatest impact on determination of the resonating frequencies (Borden & Harris, 1980). Although formant frequencies within a given speaker are not constant, there are consistent formant patterns that enable accurate perception of many English speech sounds, such as vowels, glides, and liquids. Systematic changes in the formant frequencies before or after consonants also influence listeners' identification of consonant place of articulation or voicing.

The first formant is primarily influenced by pharyngeal volume. The frequency of the first formant decreases as the place of major vocal tract constriction for a vowel or semivowel moves anteriorly, thereby enlarging the pharyngeal cavity. It also will be markedly lower if the vocal tract is constricted to some degree during production of a consonant sound preceding the vocalic segment. Additionally, the frequency of F1, as with all formants, is lowered with increasing protrusion of the lips, thereby lengthening the labial port. In fact, the effect of marked lip rounding may override the expected resonance characteristics that occur dependent upon pharyngeal volume (Kent & Read, 1992; Minifie, Hixon, & Williams, 1973).

Effect of F1 changes on the voicing contrast.

Many authors have demonstrated the importance of F1 during the transition to the postconsonantal vowel to the perception of the voicing contrast in word-initial plosives (Kewley-Port, 1982; Klatt, 1975; Lisker, 1975; Stevens & Klatt, 1974; Summerfield & Haggard, 1977). Specifically, F1 frequency at the onset of the postconsonantal vowel following voiced plosives can be expected to be lower than that following voiceless plosives. The F1 frequency during production of a stop consonant, which by definition is produced with maximal vocal tract constriction, should theoretically be at its lowest value. Therefore, the F1 frequency will always increase during the transition from the consonant to the following vowel (Kent & Read, 1992).

Voiced plosives have a well-defined first formant (F1) transition that continues after the onset of voicing. The duration of this F1 transition is approximately 40 ms or less, corresponding to the VOT range for English voiced plosives (Liberman, Delattre, Gerstman, & Cooper, 1956). The F1 transition is not as evident during production of voiceless plosives because the rapid movements of the supraglottal articulators from the plosive place of articulation to the following vowel segment are essentially complete before the onset of voice (Stevens & Klatt, 1974).

In two experiments with synthetic speech stimuli, Stevens and Klatt (1974) compared the contributions of VOT and the duration of an F1 transition following voicing onset to listeners' perceptions of a voicing distinction among plosives. VOT and F1 transition durations, from moderately fast through fast consonant-vowel transitions, were manipulated independently, and listeners' perceptions of the generated stimuli were obtained. Formant bandwidths were held constant to control the extent of the transitions. The results indicated that the perceptual boundary of the voicing contrast along the VOT dimension was not completely stable; rather, the perceptual boundary shifted with changes in the rate and duration of the F1 transition. "The presence or absence of a rapid spectral change following voice onset produces up to 15-ms change in the location of the perceived phoneme boundary as measured in terms of absolute VOT" (Stevens & Klatt, 1974, p.653). If VOT was less than 20 ms, the listeners perceived a voiced plosive. If VOT exceeded 20 ms, both VOT and F1 onset frequency influenced the listeners' decisions. In such cases in which there were relatively prolonged formant transitions following the onset of voicing, listeners were faced with conflicting cues. Some listeners assigned more weight to the relatively long VOT suggesting a voiceless plosive whereas others assigned more weight to the presence of the F1 transition suggesting a voiced plosive. Based on the results of the experiments, the authors suggested that rate and duration of the F1 transition may be more important cues than VOT to listeners' perceptions of the voicing contrast in prevocalic plosives.

Stevens ana Klatt (1974) further proposed that the perceptual importance of the F1 transition may account for longer VOT values with progressively retracted place of articulation. Slower formant transitions for alveolar and velar plosives in consonant-vowel contexts result from increased movement durations for the tongue body from consonant to vowel articulation, in comparison to the unconstrained lingual movement on transition to the vowel following release of labial plosives. Thus, for example, in order for a voiceless velar plosive to be accurately perceived, VOT must be delayed to allow the F1 transition to be completed prior to the onset of voicing (Stevens & Klatt, 1974).

Lisker (1975) questioned the results of Stevens and Klatt. Specifically, he questioned whether rate and duration were the important components of the F1 transition contributing to the perception of the voicing contrast. Lisker manipulated VOT and F1 onset in synthetically generated stimuli and obtained listener responses to the stimuli. F1 was varied according to onset frequency and the presence or absence of a transition to the vowel steady state frequency (i.e., straight versus rising F1 transitions at varying onset frequencies). His data suggested that the relevant contribution of F1 to the perception of a voiced plosive was the presence of a low F1 onset frequency, indicating maximal vocal tract constriction, and not the rate or duration of the transition as had been proposed by Stevens and Klatt (1974). Klatt (1975) suggested such low frequency energy following voicing onset should be in the region of 300 Hz and below.

Lisker's results also suggested that VOT was the stronger cue for the perception of the voicing contrast; F1 played a secondary role. Lisker noted the presence of a sharply rising F1 was most likely to occur in consonant-vowel contexts involving vowels with a high F1 steady state frequency (e.g., low vowels /ɑ,ɔ/). The effect may be negligible for high vowels /i,u/ in which the F1 steady state frequency is low. Consequently, changes in F1 onset frequency may not be apparent in these vowel contexts. Evidence of this context sensitivity argues against F1 as the primary acoustic cue for the perception of the voicing contrast (Kewley-Port, 1982; Klatt, 1975; Lisker, 1975).

Summerfield and Haggard (1977) supported and expanded upon Lisker's conclusions relative to the importance of low F1 onset frequencies to aid listeners' perceptions of voicing contrasts for word-initial plosives. In a series of experiments, the authors manipulated VOT and F1 independently in synthetic speech stimuli to determine the relative effects of each on the perception of the voicing contrast. They found that a larger VOT value was required for the perception of a voiceless stop than would ordinarily be necessary when F1 had a low onset frequency and vice versa. They identified "the perceived frequency of the first formant at the onset of voicing as the critical spectral parameter influencing the perceptual categorization of members of VOT continua" (Summerfield & Haggard, 1977, p.443).

The presence of a low F1 frequency at the onset of the vowel following a speaker's production of a voiceless plosive would result in one of two outcomes: Either the speaker would need to manipulate VOT to ensure that the listener accurately perceives the plosive, or the listener must identify the plosive when confronted with contradictory acoustic information. These contradictory acoustic cues may result in inaccurate perception of the intended plosive.

Summary.

All studies support F1 as an important cue to the perception of the voicing contrast. Various authors (Klatt, 1975; Lisker, 1975; and Summerfield & Haggard, 1977) present findings indicating that the presence of low frequency energy in the F1 transition at the onset of voicing influences the perception of a voiced prevocalic plosive. Authors disagree on whether frequency information in the F1 transition acts as a primary or secondary cue to listeners' perceptions of the voicing contrast. Stevens and Klatt (1974) argue that rate and duration of the F1 transition may be more important cues than VOT to listeners' perceptions of the voicing contrast in prevocalic plosives. However, Lisker (1975) contends that evidence of context sensitivity argues against F1 information as a primary cue.

Additional evidence is required to resolve these differences. Information from developmental studies may be particularly useful as one can observe the emerging and changing contribution of F1 to listeners' perceptions of the voicing contrast in utterances produced by children and compare the results to perceptual data for this contrast in the same utterances produced by adults.

Fundamental Frequency

Definition.

The human voice is a complex tone composed of many frequencies. The fundamental frequency (F0) is the lowest frequency of the complex tone and is perceived by the listener as the speaker's pitch (Borden & Harris, 1980). Fundamental frequency during speech is not constant; it fluctuates, providing the intonational contour of connected speech.

Developmental changes in fundamental frequency.

Developmental changes in fundamental frequency have been summarized by Kent (1976). Kent found that the periods of most rapid developmental change in fundamental frequency occur in the first four months of life, again between 1 and 3 years, and then between 13 and 17 years.

Speakers at all ages have similar fundamental frequency octave ranges (e.g., 2 1/2 to 3 octaves). However, children's fundamental frequencies are highly variable compared to those of adults. Intrasubject standard deviations for fundamental frequency decrease with age until minimum values are achieved at approximately 10 to 12 years.

Effect of F0 changes on the voicing contrast.

Fundamental frequency has been shown to function as a secondary cue to the voicing distinction among plosives

(House & Fairbanks, 1953; Klatt, 1975; Ohde, 1984; Umeda, 1981). It varies systematically with voice onset timing (Ohde, 1984). In adult speakers, postconsonantal vowel fundamental frequency is higher at voicing onset in utterances beginning with voiceless plosives than in those beginning with voiced plosives (House & Fairbanks, 1953; Klatt, 1975; Umeda, 1981). Although the difference in F0 between vowels following voiced and voiceless plosives is greatest when measured at the onset of the vowel, higher F0 values for vowels succeeding voiceless plosives compared to those following voiced plosives have been observed as far as 100 ms into the vowel (House & Fairbanks, 1953; Ohde, 1984; Umeda, 1981). Furthermore, F0 decreases faster in vowels following voicing onset for voiceless plosives compared to its decline in vowels following voiced plosives (Ohde, 1984; Umeda, 1981).

Two theories have been proposed to account for the fundamental frequency differences associated with the voicing distinction: the aerodynamic theory (Ladefoged, 1967) and the vocal cord tension theory (Halle & Stevens 1971). Ohde (1984) summarized these hypotheses.

The aerodynamic theory accounts for higher fundamental frequencies at voice onset for voiceless plosives compared to voiced plosives due to a larger pressure drop across the glottis for the voiceless stops. During the release of voiceless plosives a greater Bernoulli force than that experienced during production of voiced plosives, is caused by the high rate of airflow through the glottis. This results in a rapid closing phase of the vocal folds, causing a higher rate of vibration at voicing onset.

The vocal cord tension hypothesis asserts that the vocal folds are slack during the production of voiced plosives to facilitate voicing and stiff during production of voiceless plosives to inhibit voicing. The vertical tension in the vocal folds necessary to maintain stiffness during production of voiceless plosives results in a higher fundamental frequency at voicing onset (Ohala, 1972).

Ohde (1984) favours the latter hypothesis as the aerodynamic theory does not explain higher fundamental frequencies obtained for both voiceless unaspirated plosives as well as voiceless aspirated plosives in comparison to voiced plosives. The aerodynamic theory would predict lower fundamental frequencies for voiceless unaspirated plosives "because F0 at voicing onset would be directly relational to the degree of aspiration according to the Bernoulli principle" (Ohde, 1984, p.228). Ohde cites a number of studies with evidence suggesting that changes in the position of the laryngeal structures create tension in the larynx which influences the absolute value of F0. The production of voiceless stops, both aspirated and unaspirated, may result in increased laryngeal tension, and therefore, a higher F0 at voicing onset.

The previous discussion has been based upon data collected from adult speakers. Ohde (1985) compared VOT and fundamental frequency patterns of plosives produced by children, aged 8 and 9 years, and adults. He hypothesized that differences in fundamental frequency associated with different plosive VOT intervals would be more variable in utterances produced by children than those produced by adults. His results supported the hypothesis. Higher fundamental frequencies were observed in voiceless aspirated and unaspirated plosives as compared to voiced plosives produced by children. This effect was evident at all pitch periods under investigation with the exception of the first one.

Although the fundamental frequency difference associated with the voicing contrast (that is, higher F0 associated with production of voiceless plosives) observed in adult speech also occurred in children's speech, F0 values were substantially less stable as markers of the voicing contrast. Standard deviations for the children's data were at least twice as great as those for the adult data across all pitch periods. The greatest variability for F0 for both adults and children was observed at the first pitch period. This may partially explain the lack of statistical significance of F0 values between voiced and voiceless plosives for the first pitch period in the children's data.

Summary.

Fundamental frequency at voice onset has been shown to function as a secondary cue to the voicing distinction among plosive consonants. For adult speakers, higher absolute F0 values have been observed for voiceless plosives compared to voiced plosives at voice onset and as far as 100 ms into the postconsonantal vowel. Additionally, larger declines in F0 between vocal periods have been noted for the early stages of voice onset following voiceless plosives.

Similar trends have been noted for children. However, the F0 of children's speech is more variable than that of adult speech. Consequently, the difference in F0 at voice onset between voiced and voiceless word-initial plosives produced by children is not as great as that observed for plosive utterances produced by adults. Children do not acquire adult-like stability regarding F0 until approximately 11 years of age, whereas children may be expected to produce stable VOT values similar to those of adults by approximately 8 years of age.

Plosive Burst Characteristics

Definition.

A word-initial plosive segment is marked by an abrupt or discontinuous increase in intensity in some frequency range above 1000 Hz, occurring over a brief time interval of

20-30 ms (Stevens & Klatt, 1974). The specific peak frequencies and time intervals vary depending on the target plosive. The plosive burst conveys information to the listener about the place of articulation and voicing of the segment.

Place of plosive articulation information is conveyed on the basis of spectral energy characteristics within approximately the first 20 ms of the release burst. For example, alveolar plosives /t, d/ are characterized by either a flat spectrum, or one with most energy concentrated above 4000 Hz and another energy peak at approximately 500 Hz. In contrast, labial plosives /p, b/ are associated with a primary concentration of energy between 500 Hz and 1500 Hz, while the spectral pattern of velar plosives /k, g/ is characterized by energy concentrated between 1500 Hz and 4000 Hz (Baken, 1987, p. 372).

Plosive burst characteristics also may influence a listener's perception of the voicing distinction. Generally, the production of voiced plosives is associated with lower intraoral pressure values than is the production of voiceless plosives, and consequently results in less high frequency energy in the burst spectrum. In the case of alveolar plosives, production of the voiced /d/ may have less spectral amplitude for frequencies above 4000 Hz than the voiceless /t/. The plosive burst characteristics that may differentiate between voiced and voiceless plo ives are discussed in greater detail.

Effects of plosive burst characteristics on the voicing contrast.

Plosive burst characteristics may provide secondary cues to voicing in the acoustic signal. The amplitude of the plosive burst is the acoustic representation of intraoral pressure (Forrest & Rockman, 1988). Klatt (1975) reported that the peak intensity and the duration of the burst are greater at the release of voiceless plosives than of voiced plosives. Although burst intensity differences were not studied, Klatt argued that voiceless plosive bursts could be expected to have 3 to 6 dB greater intensity values than voiced plosive bursts given the observed differences in oral pressure. The contribution of this aspect of the burst to the perception of the voicing contrast warrants further investigation.

Fant (1973) distinguished three successive phases of the release burst of voiceless aspirated plosives: (a) the transient phase marks the increase in air pressure at the opening of the plosive occlusion; (b) the frication phase represents the turbulent air flow at the articulator constriction; and (c) the aspiration phase represents the turbulent air flow at the level of the vocal folds.

Voiced plosive bursts are marked only by the transient and frication phases. There is no aspiration phase, presumably because the vocal folds are held in an adducted position to allow vibration; consequently, "the glottal aperture is not large enough for sufficient flow to produce a turbulent source of aspiration" (Revoile, Pickett, Holden-Pitt, Talkin, & Brandt, 1987, p.7).

Summary of Acoustic Cues Related to the Voicing Contrast

In the previous discussion, four acoustic cues thought to contribute to listeners' perception of the voicing contrast were defined and relevant literature outlining changes in these cues as a function of the voicing contrast in utterances produced by children and adults were reviewed. Primary points that should be taken away from this discussion are as follows:

1. VOT has historically been considered the most salient cue to listeners' perceptions of the voicing contrast. VOT may be a stronger, more consistent cue to the perception of voicing as speaker age increases due to the highly variable VOT values in utterances produced by young children.
2. F1 has also been considered an important cue to listeners' perceptions of voicing. However, the literature provides contradictory information regarding which aspects of the F1 transition (e.g., rate or extent of the transition, duration, onset frequency) influence perception, and the importance of these features to listeners' perceptions of the voicing contrast. As with VOT, F1

frequency information may be a stronger, more consistent cue to the perception of voicing as speaker age increases due to high variability of formant values in utterances produced by children.

3. F0 changes associated with the voicing contrast provide secondary cues to help the listener accurately perceive the voicing contrast. Due to the high variability in F0 until approximately 12 years of age, F0 may be a more consistent cue to the perception of voicing as speaker age increases.

4. Plosive burst amplitude characteristics associated with the voicing contrast change passively as a result of the action of the vocal folds and consequent differences in intraoral pressure during production of voiced and voiceless plosives. These characteristics probably provide secondary information to listeners' perceptions of voicing and may be overridden by changes in more salient cues, such as VOT or F1 onset frequency. However, this aspect of the voicing contrast has not been as extensively studied as have the other acoustic variables.

5. These acoustic cues are thought to be highly interdependent and can be modified by a variety of contextual factors. Changes in one of the parameters may affect the others, thereby affecting listeners' perceptions of the voicing contrast. However, the exact contribution of any one factor or combination of factors to the perception of the voicing contrast in word-initial plosives merits

further study.

It is likely that there is a complex interaction of factors that impact upon the listener's perception of the voicing contrast. The listener needs to evaluate the relative importance of all factors to identify a sound. If the cues are ambiguous, the listener may incorrectly perceive the target sound. If the speaker is a child, the listener's task may be more difficult. The immature neurological system of the child results in greater variability of VOT values. Similarly, the child's control over the source of other acoustic cues, such as F1 and F0, will vary. Thus, the listener may have to identify the target sound in the presence of conflicting cues. Which of these cues is most important to the listener's perception of the voicing contrast? Few studies of the voicing contrast produced by children have included a perceptual validation component.

Studies of Voicing Including Perceptual Validation Tasks

Menyuk and Klatt (1974) studied VOT in consonant clusters produced by children and adults. In addition to the VOT measures, 3 listeners trained in phonetic transcription transcribed the consonant clusters contained within isolated words and sentences from audiotape recordings.

Most consonant clusters were perceived to be produced correctly. The acoustical results did not indicate a unique temporal boundary for the perception of the voicing distinction. There was significant variability in and overlap among VOT values across voicing categories, particularly for the children's data, even when the perceived and intended productions were accurately matched. This suggested that other factors, such as information from formant transitions or onsonantal context, may have influenced the listeners' perceptions of the voicing contrast in consonant clusters. The authors recommended that future research be directed towards spectrographic analyses of various aspects of children's speech and correlation of such analyses of speech production with information regarding speech perception.

There are some interesting studies involving phonologically-disordered children's productions of VOT and the relationship of VOT to listeners' perceptions of the voicing contrast. Catts and Jensen (1983) investigated the voicing contrast in word-initial and word-final plosives produced by 9 phonologically-disordered and 9 normally-developing children. Measurements of VOT, vowel duration, and consonant closure were made. Additionally, 9 listeners trained in phonetic transcription transcribed audiotape recordings of the children's productions of minimal pair words differing only on the voicing feature. A percentage of voicing error was calculated for each subject group by dividing the number of listeners' transcriptions indicating a voicing error by the total number of listeners' transcriptions.

Regarding word-initial plosives, the results indicated that the normally-developing children produced distinctive VOT values differentiating voiced from voiceless plosives. The phonologically-disordered children displayed one of two patterns: Either they failed to produce contrastive VOT values so that all VOT values fell within the short-lag range or they produced exaggerated VOT contrasts in which voiceless plosives were marked by excessively long VOT values. The perceptual analysis indicated significantly more voicing errors identified for the phonologically- disordered children than for the normally-developing children. However, not all of the perceived voicing errors for words produced by phonologically-disordered children could be accounted for solely on the basis of VOT. The contribution of other acoustic cues warranted further investigation.

Maxwell and Weismer (1982) studied the voicing contrast produced by one phonologically-disordered child. They found a statistically significant difference in VOT for the child's attempts to produce voiced sounds compared to voiceless sounds, even though a panel of trained listeners identified all of the child's plosives as voiced. They concluded that for some misarticulating children, the assumption that they have limited phonological knowledge regarding the voicing contrast may be incorrect. In actuality, these children's attempts to produce a voicing contrast may not be perceptually salient.

Finally, Forrest and Rockman (1988) compared the relationship between various acoustic cues and normal listeners' perceptions of the voicing contrast for word-initial plosives produced by phonologically-disordered children. Three phonologically-disordered boys were required to spontaneously produce a list of eighteen monosyllabic words. A subset of the words was elicited in a carrier phrase. Fifteen normal listeners were randomly assigned to one of three groups corresponding to each of the speakers. The listeners were asked to judge the word-initial plosive of the tape-recorded target words for the presence or absence of voicing along a 7-point scale.

[illegible] capacity of VOT to predict perceived voicing was not high. Consequently, the authors investigated the influence of a matrix of acoustic cues, including F1 onset frequency, F0 onset frequency, burst and aspiration amplitude, to listeners' perceptions of word-initial voicing as produced by phonologically disordered children. The authors concluded that no single acoustic cue could account for listeners' perceptions of the voicing contrast for either normally-developing or phonologically-disordered speakers. Some combination of cues accounted for the perceived voicing of approximately half of the words that were not differentiated by VOT alone (Forrest & Rockman, 1988).

Further investigation is warranted to determine the complex relationship between acoustic correlates of the voicing contrast and listeners' perceptions of that contrast. Many of the aforementioned studies focused on the effects of VOT on the perception of the voicing contrast produced by children. However, as has been shown, VOT alone may not be an adequate predictor of the perception of the voicing contrast. Further study of spectral characteristics thought to be associated with the voicing contrast, such as F1 transitions, is a frequent recommendation. Unfortunately, the difficulties inherent in measurement of the spectral characteristics of children's speech have impeded the development of research in this area. Some of these difficulties are described below.

Limitations of Acoustic Analysis

Acoustic measurement of voices of higher fundamental frequencies, particularly children's voices, is subject to substantial measurement error. Measurement error may arise from the peculiarities of the speech signal produced by children, limitations of instrumentation adapted from adult speech measurement (Kent, 1976), and examiner error arising from incorrect identification of the acoustic parameters within the speech signal (Baken, 1987; Kent & Read, 1992).

The idiosyncrasies resulting from immature neuromotor control in a child's developing motor-speech system may obscure the acoustic details of the speech signal. For example, children frequently use inappropriate nasalization which results in unexpected resonances and antiresonances. Additionally, children's voices may be breathy or hoarse, which introduces noise into the speech signal, thereby obscuring other acoustic details (Kent, 1976).

Measurement of children's formant frequencies is particularly difficult. Formant frequency estimation becomes less accurate as fundamental frequency increases. The widely spaced harmonics of the vocal source do not always approximate resonant frequencies of the vocal tract sufficiently to convey spectral information necessary for vowel identification. This may result in loss of information in the spectrographic analysis (Baken, 1987).

Measurements of spectral characteristics can be made by visual inspection of spectrographic displays of the speech signal in analog or digital form, or by means of various other computerized (digital) analyses such as linear prediction coding (LPC). Both measurement procedures require that the examiner visually identify the formant frequencies and the points from which they will be measured. Consequently, both measurement techniques are susceptible to sampling error resulting from the unsuitability of conventional analyzing bandwidths for voices with high fundamental frequencies (Kent, 1976) and variable source characteristics.

Specifically, the widely spaced harmonics associated with high fundamental frequencies result in poor resolution of the resonating frequencies when a typical analyzing bandwidth of 300 Hz is used. There is an increased risk for formant-harmonic interaction; the investigator, or computer, may incorrectly judge a strong harmonic to be the centre frequency of a formant, or may not resolve the formant at all (Kent, 1976). In fact, based on data presented by Monsen and Engebretson (1983), measurement error of the first two formants may be as great as +/- 60 Hz.

A few procedures may be employed to improve the accuracy of acoustic measurements of speech signals produced by speakers with high fundamental frequencies. For example, wider analyzing bandwidths may improve the resolution of the child's vocal tract resonance characteristics (Kent, 1976). Wide-band analysis emphasizes formant information whereas narrow-band analysis highlights the harmonics. Generally, the analyzing bandwidth should be two to three times larger than the speaker's fundamental frequency to encompass at least two harmonics, thereby making the formant resonances maximally visible (Keller, 1992; Kent & Read, 1992). However, although current technology makes such bandwidth analysis possible, it does not eliminate the difficulty of measuring spectral characteristics of children's speech.

Caution in interpreting such data from acoustic measurements remains necessary.

Spectral Normalization

Kent and Forner (1979) discussed the difficulties of equating formant frequency measures of vowels with their appropriate phonetic equivalents. Formant frequency measurements for any one speaker display considerable variability. Formant frequency values may vary greatly within a vowel category, and overlap across vowel categories.

Added to the difficulty of making within-speaker comparisons is the difficulty of comparing spectral information produced by speakers with different vocal tract sizes and structure. Differences in vocal tract size and structure differentially influence the relationship of formant frequencies within and across vowels, even though the vowels retain their perceptual identity (Hodge, 1989). Inferences regarding inter-speaker differences in spectral information cannot be made without applying a spectral normalization procedure.

A variety of approaches has been proposed to normalize spectral differences resulting from different vocal tract sizes. One such method is the Bark Transformation described by Syrdal and Gopal (1986). As Hodge (1989) summarized, the Bark transformation equation converts frequency (in kHz) to a critical band, or Bark, scale approximating the proposed human auditory representation of frequency. Conversion of spectral values expressed in kHz to the Bark scale is not sufficient to normalize frequency values for vocal tract size differences; rather, values transformed to the Bark scale need to be expressed as a difference of one value from another (e.g., the difference between F1 and F0 in Bark).

The transformation of spectral measures to the Bark scale allows direct comparisons of spectral information among various speakers. As part of her study, Hodge (1989) investigated whether the Bark transformation eliminated absolute spectral differences resulting from vocal tract size differences represented by speakers from early childhood to adulthood. Estimates of the fundamental frequency and the first three formants (originally expressed in kHz) of 5 vowels produced by the speakers were converted to the Bark scale and three Bark differences for each vowel were calculated (i.e., F1 - F0, F2 - F1, F3 - F2). The author found that, with some limitations, the Bark transformation successfully normalized spectral differences resulting from different vocal tract sizes across speakers.

The advantage of using the Bark transformation is that each spectral value can be expressed in Bark values which allows for comparisons of differences in individual formants. The disadvantage relates to the manner in which the Bark transformation handles the data. The Bark scale represents linear functions of frequency in Hz in the low-frequency ranges, transitional functions in the mid-frequency ranges, and logarithmic functions in the high-frequency ranges (Miller, 1989, p. 2116). Specifically, Bark values for frequencies below 500 Hz are linearly related to Hertz; above 1000 Hz, the relationship of Bark to Hertz is expressed logarithmically (Nearey, 1989). Differences between logarithmic or Bark transformations are minimal in the mid- and high-frequency ranges. However, the apparent magnitude of changes in the low-frequency ranges may be larger using the Bark transformation than could be expected to occur using a logarithmic transformation (Miller, 1989; Nearey, 1989). This may impact on the results of the present investigation as F0 frequencies for all subjects and F1 frequencies for adult subjects may fall within this low-frequency range.

In the present study, two normalization procedures have been employed in the measurement of first formant frequency and fundamental frequency for the vowel / /. Specifically, F0 and F1 values have been converted to both Bark and logarithmic scales. The decision as to which scaled values to include in the statistical comparisons has been empirically determined.

Purpose

The purpose of this study was to examine the combined and individual effects of selected acoustic cues on listeners' perceptions of the voicing contrast in word-initial alveolar plosives produced by speakers in four age groups who were judged to have age-appropriate speech and vocal quality. This study was intended to explore three areas. Firstly, this study attempted to document differences in knowledgeable listeners' abilities to accurately perceive voicing contrasts of word-initial alveolar plosives produced by speakers of various ages and varying levels of neurological maturity. Secondly, this study attempted to document differences that may exist in the selected acoustic cues between perceived voicing condition and across age group. Finally, this study attempted to determine the relative importance of the acoustic cues associated with the perception of word-initial voiced and voiceless plosives for each age group.

The acoustic cues investigated included (a) VOT measured in milliseconds (ms); (b) first formant frequency (F1) at the onset of voicing following the plosive segment, measured in Hertz (Hz) and converted to Bark and natural logarithm scales; (c) fundamental frequency (F0) at the onset of the vowel following the plosive segment, measured in Hertz (Hz) and converted to Bark and natural logarithm scales; and (d) plosive burst amplitude expressed as a ratio of the amplitude of the plosive burst to the maximum amplitude of the following vowel, measured in Volts (V). These acoustic cues were measured from speech waveforms recorded live, stored on audiotape, and digitized using the CSpeech (Milenkovic, 1989) microcomputer software program for speech analysis. Values of the spectral measures were originally made in Hertz, then converted to the Bark and natural logarithm scales to allow statistical comparisons among speakers with varying vocal tract sizes and structures.

Normal speakers at four age levels acted as speaker subjects for this study: 2 years 6 months to 3 years (referred to as 2.5 year olds); 4 years 6 months to 5 years (referred to as 4.5 year olds); 10 years to 11 years (referred to as 10 year olds); and adults ranging in age from 24 to 38 years. The age groups were selected to reflect the progression of acquisition of the voicing contrast from emergence to mastery.

Specifically, the following questions were asked:

(1) Are there differences in listeners' perceptions of the voicing contrast in word-initial alveolar plosives produced by speakers of various ages?

(a) Are there significant differences in the frequency of correct listener judgments for identification of word-initial alveolar plosives, differing on the voicing feature, produced by 2.5 year old children, 4.5 year old children, 10 year old children, and adults? Correct listener judgments were defined as each occurrence in which the word-initial alveolar plosive was identified by a listener as it was intended by the speaker with respect to the voicing feature (e.g., the speaker's production of /t/ was identified by the listener as /t/).

(b) Are there significant differences in the frequency of incorrect listener judgments for identification of word-initial alveolar plosives, differing on the voicing feature, produced by 2.5 year old children, 4.5 year old children, 10 year old children, and adults? Incorrect judgments were defined as each occurrence in which the word-initial alveolar plosive was identified by a listener as opposite to the way the speaker intended with respect to the voicing feature (e.g., the speaker's production of /t/ was identified by the listener as /d/).

(c) Are there significant differences in the frequency of ambiguous judgments for identification of word-initial alveolar plosives, differing on the voicing feature, produced by 2.5 year old children, 4.5 year old children, 10 year old children, and adults? Ambiguous judgments were defined as each occurrence in which the word-initial alveolar plosive was identified by the listener as a questionable /t/ or questionable /d/.

(2) Are there significant differences in VOT among perceptually-validated voiced and voiceless word-initial

alveolar plosives produced by 2.5 year old children, 4.5 year old children, 10 year old children, and adults? A perceptually-validated item was defined as a token that has been correctly identified by four out of five knowledgeable adult listeners.

(3) Are there significant differences in F1 at the onset of the vowel following the plosive burst among perceptually-validated voiced and voiceless word initial alveolar plosives produced by 2.5 year old children, 4.5 year old children, 10 year old children, and adults?

(4) Are there significant differences in F0 at the onset of the vowel following the plosive burst among perceptually-validated voiced and voiceless word-initial alveolar plosives produced by 2.5 year old children, 4.5 year old children, 10 year old children, and adults?

(5) Are there significant differences burst amplitude among perceptually-validated voiced and voiceless word-initial alveolar plosives produced by 2.5 year old children, 4.5 year old children, 10 year old children, and adults?

(6) What combination of these 4 possible acoustic cues, namely VOT, first formant frequency at the onset of voicing, fundamental frequency at the onset of voicing, and plosive burst amplitude, best predicts listeners' identification of word-initial voiced and voiceless plosives produced by normal children and adults?

The following hypotheses regarding these questions were made:

(1a) The frequency of correct listener judgments for identification of voiced and voiceless word-initial alveolar plosives will increase as speaker age increases.

(1b) The frequency of incorrect listener judgments for identification of voiced and voiceless word-initial alveolar plosives will decrease as speaker age increases, with the greatest number of incorrect judgments occurring for the youngest speaker group.

(1c) The frequency of ambiguous judgments for identification of voiced and voiceless word-initial alveolar plosives will decrease as speaker age increases, with the greatest number of ambiguous judgments occurring for the youngest speakers.

(2a) There will be a significant main effect on voicing condition for VOT, with longer mean VOT measures for voiceless alveolar plosives than for voiced alveolar plosives.

(2b) There will be an interaction effect between age and voicing condition in terms of VOT. Specifically, VOT values for voiceless word-initial alveolar plosives produced by the three oldest speaker groups will be significantly longer than those produced by the youngest group.

(3a) There will be a significant main effect on voicing condition for the postconsonantal vowel's F1 onset frequency, with consistently lower mean F1 onset frequency values for voiced alveolar plosives than for voiceless alveolar plosives.

(3b) There will be an interaction effect between age and voicing condition on the postconsonantal vowel's F1 onset frequency. Specifically, differences in the postconsonantal vowel's F1 onset frequency between voiced and voiceless word-initial alveolar plosives will be significant only for the two oldest groups.

(4a) There will be a significant main effect on voicing condition for F0 at the onset of the postconsonantal vowel, with consistently higher mean fundamental frequency values for voiceless alveolar plosives than for voiced alveolar plosives.

(4b) There will be an interaction effect between age and voicing condition on mean F0 values. Specifically, differences in the postconsonantal vowel's F0 onset frequency between voiced and voiceless word-initial alveolar plosives will be significant only for the two oldest groups.

(5a) There will be a significant main effect on voicing condition for burst amplitude, with the mean amplitude measures for voiceless alveolar plosives consistently higher than those for voiced alveolar plosives.

(5b) There will be no significant difference in burst amplitude between age groups.

(6) The capacity of a combination of acoustic variables (including VOT, F1 - F0 differences at onset of the postconsonantal vowel, and plosive burst amplitude) to predict listeners' perceptions of the voicing contrast for word-initial /t/ and /d/ will improve with speaker age, and more cues will contribute to this prediction as speaker age increases.

Method

Design

This study employed a combined group comparative and correlational design. The perceptual and acoustical data were compared between voicing conditions and among subject groups, then the acoustical data were related to the perceptual data by means of a simple discriminant function analysis.

The first experimental question, which considered the differences in accuracy of listeners' perceptions of word-initial alveolar plosives produced by speakers of various ages, was answered within a comparative mixed 2-factor (4 x 2) univariate design with repeated measures on the factor of voicing condition. The independent variables included (1) speaker age group with four levels (2.5 - 3 year old children, 4.5 - 5 year old children, 10 - 11 year old children, and adults), and (2) voicing condition with two levels (voiced word-initial alveolar plosives and voiceless word-initial alveolar plosives). The dependent variable consisted of listeners' classifications of the word-initial plosives in the words spoken by the various speaker-subjects.

Experimental questions 2 through 5, which considered differences that may exist in the various acoustic variables relative to voicing condition and speaker age, were answered within a comparative 2-factor (4 x 2) multivariate design with repeated measures on the factor of voicing condition. One perceptually-validated token for each plosive (/t/ and /d/) produced by each speaker-subject was randomly selected for inclusion in the multivariate analysis of variance (MANOVA). As for the analysis of the perceptual data, the independent variables included (1) speaker age group with its four levels and (2) voicing condition with its two levels. The four dependent variables included VOT, F1 frequency at voicing onset of the vowel, F0 at voicing onset of the vowel, and plosive burst amplitude.

The remaining experimental question, which explored the relative contribution of the acoustic cues associated with the voicing contrast to listeners' perceptions of word-initial alveolar plosives, was addressed through a descriptive correlational design using a simple discriminant function analysis. The acoustic variables used in the MANOVA analysis became the predictor variables in the discriminant function analysis. The criterion (or dependent) variables were the listeners' identifications of the word-initial alveolar plosives as either /t/ or /d/. The perceptually-validated tokens that had been randomly selected for inclusion in the MANOVA analysis were also used for the discriminant function analysis.

A sample size of 20 speaker-subjects per age group was calculated _a priori_, based on estimations of anticipated critical effect size and power level of the study. (The equation used followed procedures for calculation of sample size appropriate for a multiple linear regression design.) Specifically, in a regression equation where R^2 =.20 and f^2 =.80, a total subject pool of at least 53 subjects was required (α= .05, β= .20). (Appendix A -- Sample size calculation).

Subjects

A total of 80 normally-speaking children and adults served as speaker-subjects. Twenty children were selected from each of three age groups: 2 years 6 months to 3 years (M = 2 years 9 months), 4 years 6 months to 5 years (M = 4 years 9 months), and 10 years to 11 years (M = 10 years 7 months). Additionally, 20 adults were selected for participation in the study. Adults ranged in age from 24 years to 38 years (M = 30 years). Originally, the investigator had proposed to include only child speakers, hypothesizing that the 10 year old children would display speech development and stability comparable to that of adults. However, because the literature suggests that adult-like stability regarding F0 and F1 is usually but not always acquired by 11 years of age (Kent, 1976), it was felt that the use of an adult group was warranted. Equal numbers of males and females were included. All speakers met the following criteria to serve as speaker-subjects:

(1) Hearing acuity, as tested by means of standard

audiologic screening procedures, was within normal limits, bilaterally.

A qualified speech-language pathologist screened subjects' hearing acuities at 20 dB HL for the frequencies of 500, 1000, 2000 and 4000 Hz (ANSI, 1970). The auditory stimuli were presented under headphones to the right and left ears in a quiet environment. The hearing screening was performed using a portable audiometer in a quiet location at the various testing sites. For the purposes of this study, a subject was considered to demonstrate hearing acuity within normal limits if the standard hearing screening procedure was passed for both ears.

(2) Visual acuity, with corrective lenses if necessary, was within normal limits according to parents' or adult subjects' reports.

(3) Structure and function of the oral-peripheral mechanism were within normal limits.

Subjects were given a cursory oral-peripheral examination by a qualified speech-language pathologist. Additionally, interviews with parents, daycare personnel, or adult-subjects were conducted to ensure that there was no history of abnormalities of speech mechanism structure or function. Potential subjects exhibiting gross structural abnormalities of the orofacial region which might impair articulation (e.g., repaired cleft palate or severe dental malocclusion) were not included in the study.

(4) English was each subject's first language (as VOT values differ across languages).

(5) Vocal quality, as assessed by a qualified speech-language pathologist using the Wilson Voice Profile Rating System (Wilson, 1971), was within normal limits.

Unusual vocal quality, such as excessive breathiness, may adversely affect VOT; consequently, speakers who were judged to demonstrate an unusual vocal quality were excluded from the study.

(6) Speech development was within normal limits.

The principal investigator, a speech-language pathologist, administered the articulation tests and judged the spontaneous speech samples. Normalcy of speech production by children younger than 3 years was based on judgment of the child's speech during a brief spontaneous speech sample and administration of the articulation screening task of the Preschool Language Scale - Revised (Steiner & Pond, 1979). Normalcy of speech production by subjects aged 3 years to 5 years was assessed through administration of the 50-item screening portion of the Templin-Darley Test of Articulation (Templin & Darley, 1968) and judgment of a brief spontaneous speech sample. Age-appropriate speech production in each potential subject was based on the subject's attaining a score at or above the cutoff for his age on the articulation screening test, and the investigator's subjective analysis of the speech sample.

For subjects aged 9 years and older, the investigator judged a brief speech sample to determine normalcy of speech production.

(7) Academic, social and emotional development, evaluated on the basis of parent report and observation of the subject's behaviour, were within normal limits.

Formal cognitive testing was not conducted. Children and adult subjects included in the study did not display any noticeable cognitive, social-emotional, or academic difficulties nor was there any known history of substantial medical or psychiatric involvement, or special education requirements.

Judgments of normal cognitive, academic, social and emotional development of the children were based on verbal reports by parent or daycare staff. Adult subjects were all known to the investigator. They were engaged in a variety of occupations, including business or health professions, or university studies.

Each speaker-subject participated in an experimental session that lasted approximately one hour and occurred either in the subject's home or, in the case of many of the children, in a daycare setting. Both the screening and testing procedures were administered within the same session. Testing procedures were not administered to any potential subject who did not pass any component of the screening procedures. The experimental session was concluded once any portion of the screening procedures was not passed. In this case, the subject (or the subject's parents) was (were) notified of the results of the screening procedures and referrals to physicians or appropriate treatment agencies were made if requested. The numbers of potential subjects that were excluded from participation in the study were as follows: four 2.5 year old children, eight 4.5 year old children, three 10 year old children, and 2 adults (Appendix B -- Summary of potential subjects who did not pass the screening procedures).

Subjects (or, in the case of children, their legal guardians) were notified of the purpose and procedures of the study in advance by means of information forms, and informed consent was obtained from all participants.

Procedure

Production Task

Minimal pair consonant-vowel-consonant (CVC) words representing the voicing contrast in lingua-alveolar plosives (i.e., /t/ and /d/) in the initial position followed by the vowel /ɑ/ (as in /dɑt/ -- "dot", or /tɑt/ -- "tot") were used (Appendix C -- List of words included in the experimental task). Of the three places of plosive articulation possible in English (bilabial, alveolar, and velar), the alveolar place was chosen for this analysis as it has the greatest frequency of occurrence (Shriberg &

Kent, 1982) and avoids potential problems of velar fronting and multiple stop bursts evidenced in younger children's normal productions of velar plosives (Shriberg & Kwiatkowski, 1980; Forrest & Rockman, 1988). The postconsonantal vowel context /ɑ/ was chosen because previous research indicated that low first formant onset frequencies can be tracked reliably for this vowel (Forrest & Rockman, 1988).

A head-mounted microphone (Shure Model SM 10A, low impedance, unidirectional, dynamic) was used to transduce speaker-subjects' spontaneous single-word utterances spoken at a comfortable loudness level. The microphone was positioned at the corner of the speaker-subject's mouth. The microphone was coupled to a 2-track portable reel-to-reel audio tape recorder (Nagra 4.2). Seven-inch magnetic audio recording tapes (3M 808 low print) were used to store the microphone signal on one track of the recorder.

Each speaker-subject was shown a picture and asked to complete the carrier phrase "This is a _____." The carrier phrase was spoken by the examiner; the speaker-subject was trained to complete the sentence frame. The speakers were familiarized with the words associated with each picture card and asked to rehearse aloud the label for each picture card prior to beginning the test procedures. Additional words other than the test words were included in the production task to keep the speakers interested in the task, and to allow easier randomization of the test words (thereby avoiding multiple sequential repetitions of any test word).

Occasionally, children within the two youngest age groups forgot the target word or were reticent to respond. When this occurred, the investigator provided a verbal model for the child to repeat; opportunity for the word to be produced again spontaneously by the child was provided later in the task.

Each child produced six tokens of each of two test words (i.e., "tot" and "dot"). The principal investigator selected only five tokens of each word produced spontaneously by each speaker to present to listeners for perceptual validation. The extra response was collected in case any one had to be discarded due to such factors as interfering background noise or a low-intensity response level by the speaker-subject (Catts & Jensen, 1983). A total of 880 tokens were included in both the perceptual validation analysis and the acoustic analysis of the words spoken (i.e., 80 speakers x 5 repetitions x 2 test words). A subset of 80 of these tokens (10% of the data from each age group) were presented twice in the perceptual task to allow for calculation of intra-listener reliability rates.

The production task format, that is, the carrier phrase spoken by the examiner and a subject's spontaneous utterance of a word to complete the carrier phrase, served several functions. Contextual variables which could affect the acoustic variables of interest could be better controlled, thereby improving the chances that any differences in the values of the acoustic variables between groups or within speaker-subjects could be attributed primarily to differences between the word-initial phoneme in the words spoken. For example, use of the sentence frame to elicit the target was designed to control varying inflectional markers. Measurement of the target phoneme produced in single words minimized the effects of phonetic context. It was anticipated that the spontaneous production task would yield utterances with VOT values that more closely approximate actual VOT values for naturally-produced single words than for words repeated in an immediate imitation task. Additionally, this task allowed the principal investigator some control of rate and pitch fluctuations, particularly for the older speaker-subjects who demonstrated more knowledge of conversational rules.

It is acknowledged, however, that the experimental task was highly artificial. Therefore, any effect of the acoustic variables on the perception of the voicing contrast identified in this study can be interpreted only within its experimental context and not with respect to spontaneous connected speech.

Acoustical Analyses

The acoustical analyses of the speaker-subjects' recorded speech were performed by computer using a low pass (antialiasing) filter (Frequency Devices 901), a 12-bit A/D board (Data Translation 2821D) and CSpeech (Milenkovic, 1989), a commercially-available digital speech waveform analysis system written in Turbo Pascal and designed to operate on an IBM - PC/AT microcomputer.

CSpeech was chosen because it afforded the experimenter a number of desirable waveform analysis features. It allows high sampling rates (up to 40 kHz per data channel), and can perform linear predictive coefficient (LPC) and Fast Fourier Transform (FFT) analyses. Values of time (in ms), frequency (in Hz), and amplitude (in dB or RMS) can be displayed for specified points on the waveform in both amplitude x frequency spectral displays and waveform displays. Finally, it is compatible with the PERCEPT software operating with the Canadian Speech Research Environment (CSRE) (Jamieson, 1989) for ease in programming presentation of the acoustical data for perceptual analysis.

For every speaker-subject, the audio tape recording of each production of a test word was low-passed filtered at 10.5 kHz and digitized at 26 kHz, or 2.56 times 10 kHz, the highest expected frequency in the speech samples (Enochson, 1986). This expectation was based on the work of Hodge (1989) who estimated that 10 kHz will be the highest frequency within the spectral display of the plosive burst for /t/ produced by 4.5 year old children. Therefore, the low-pass filter corner frequency of 10.5 kHz accommodated

the frequency range of interest in this investigation but precluded aliasing (a pitfall in digital acoustical analyses) by ensuring that acoustical energy greater than 10.5 kHz would be attenuated by 15.6 dB at 13 kHz, the Nyquist frequency for a 26 kHz sampling rate.

Voice onset time.

Voice onset time (in ms) of the digitized signals was measured as the time elapsed between the release of a word initial plosive burst and the zero crossing of the first positive-going, large amplitude, well-formed repetitive wave shape associated with the following vowel. This interval was designated by the investigator using cursors at the specified points on the visual display of the target waveform; the duration of the interval (VOT) was automatically calculated by the CSpeech software (Figure 1).

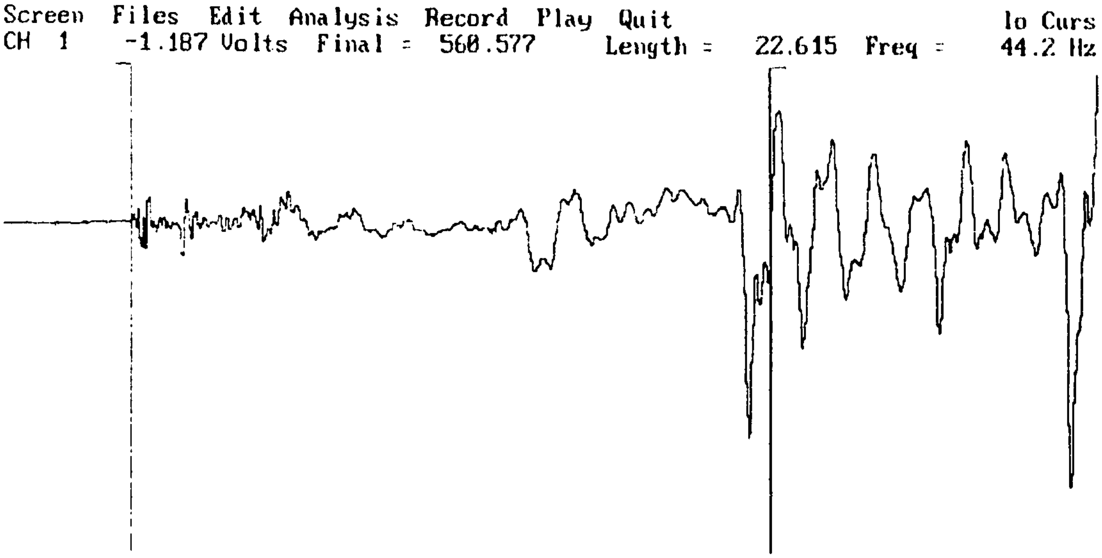
First formant onset frequency.

First formant onset frequency (in Hz) was calculated on a portion of the first pitch period of the postconsonantal vowel. F1 frequency values were obtained with reference to FFT and LPC analyses of just less than or equal to 1/2 of the first pitch period (i.e., effective analysis bandwidth = 2 x the speaker's fundamental frequency). This procedure was guided by Milenkovic's (1989) and Keller's (1992) assertions that isolating a smaller than usual analysis

Figure 1

EXAMPLE OF VOT MEASUREMENT FOR /d/

ON A WAVEFORM OF /dɑt/ SPOKEN BY AN ADULT MALE



```
CH 1    -1.187 Volts  Final =  560.577    Length =   22.615  Freq =   44.2 Hz
```

-- cursors are placed from the onset of the plosive burst to the zero crossing of the first pitch period of the vowel.

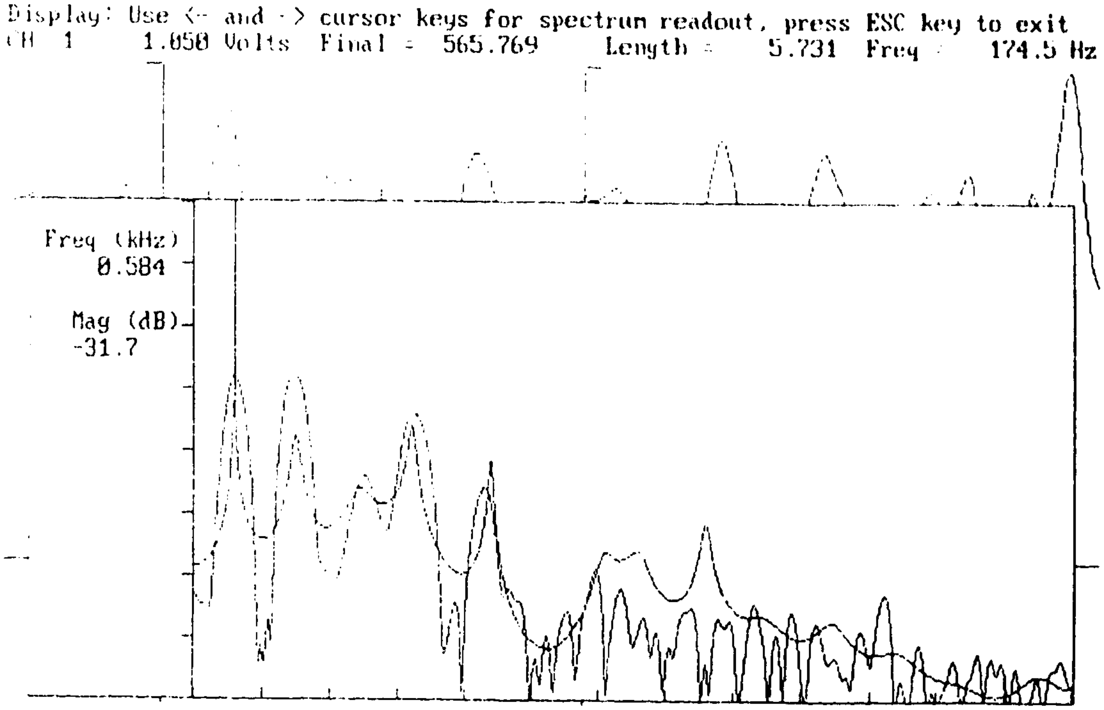
window will simulate wide-band spectral analysis necessary for optimal resolution of formant measurements of speech produced by subjects with higher fundamental frequencies.

Measurement of F1 onset frequency began by isolating the first well-formed, large amplitude pitch period indicating the onset of the vowel. A portion that was just less than or equal to 1/2 that of the full pitch period was further isolated beginning at the zero crossing of the first positive-going, large-amplitude peak, and FFT and LPC spectral displays were generated. The number of LPC coefficients determined by the CSpeech software program to calculate F1 frequency values for each time frame of interest totalled 28 (based on a calculation of sampling frequency + 2, i.e., 26 kHz + 2 = 28). The results of both FFT and LPC amplitude x frequency spectral displays for the interval selected could be superimposed on the computer screen. With visual reference to both displays, the investigator placed the CSpeech cursor on the peak in the FFT spectrum thought to correspond to the first formant. The frequency value on the spectrum corresponding to that cursor position was automatically calculated and displayed by the CSpeech software (Figure 2). Although both methods were used to aid the investigator in the measurement of F1, only the values obtained from the FFT analysis were included in the statistical comparisons. These frequency values later were converted to Bark and natural logarithm values

Figure 2

EXAMPLE OF F1 MEASUREMENT FROM FFT AND LPC SPECTRA

OF A WAVEFORM OF /ɑ/ IN /dɑt/ SPOKEN BY AN ADULT MALE



```
Display: Use <- and -> cursor keys for spectrum readout, press ESC key to exit
CH 1     1.058 Volts  Final =  565.769    Length =   5.731  Freq =   174.5 Hz
```

--FFT and LPC spectra were calculated for an analyzing window equivalent to 1/2 pitch period (i.e., 5.731 ms). The cursor is placed on the FFT peak nearest to F1.

via spreadsheet software (Excel 1.5) on a Macintosh computer to allow statistical comparisons across age groups (Appendix D -- Bark Transformation Equation).

Fundamental frequency.

A fundamental frequency (F0) value (in Hz) for each digitized signal was obtained for the first pitch period of the postconsonantal vowel. The CSpeech cursors were placed at the first positive-going point of the peaks marking the beginning and end of the first pitch period. On the basis of the duration of this period, F0 was automatically estimated by the CSpeech software (Figure 3). These frequency values were converted to Bark and natural logarithm scales via spreadsheet software (Excel 1.5) on a Macintosh computer to allow for statistical comparisons.

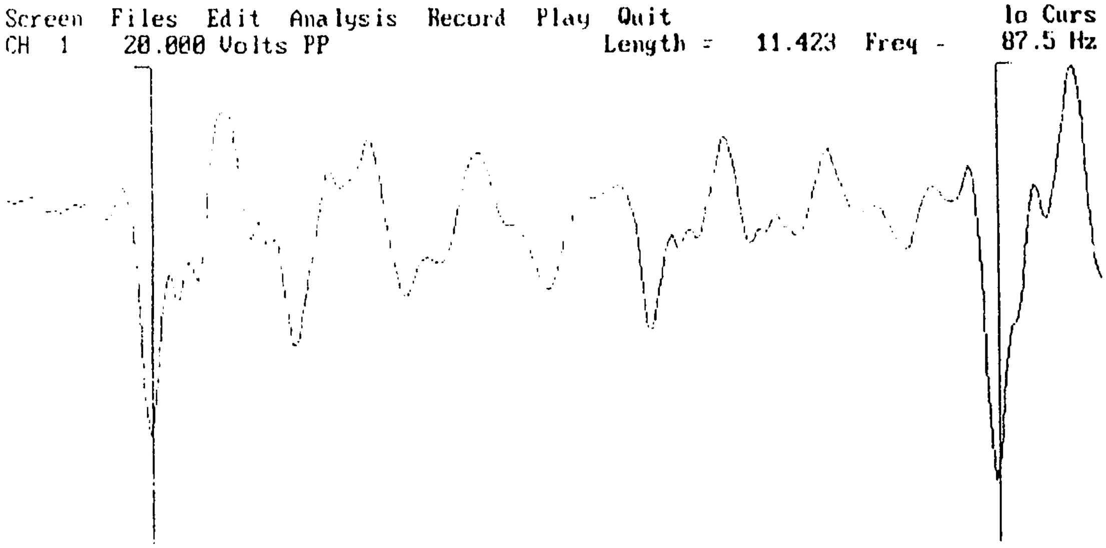
Plosive burst amplitude.

The amplitude of the plosive burst was measured in relation to the point of greatest amplitude of the following vowel. The investigator isolated the plosive burst energy in the time x amplitude display of the waveform and placed the cursor on the highest peak. The amplitude (in Volts) corresponding to the designated peak was automatically calculated by the CSpeech software program (Figure 4).

Figure 3

EXAMPLE OF F0 MEASUREMENT FOR ONE PITCH PERIOD

OF A WAVEFORM OF /ɑ/ IN /dɑt/ SPOKEN BY AN ADULT MALE



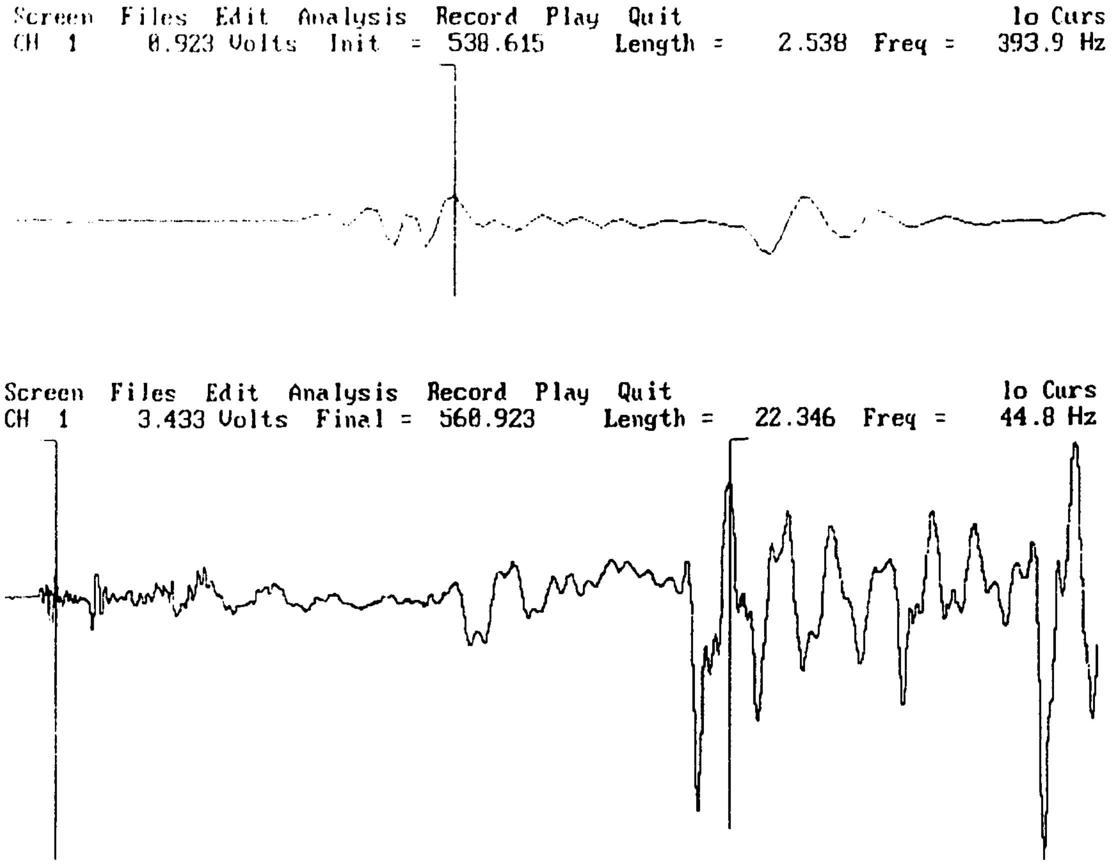
```
CH 1   20.000 Volts PP                 Length =   11.423  Freq -   87.5 Hz
```

--cursors are placed at the beginning and end of the first pitch period of the postconsonantal vowel. The fundamental frequency in Hertz estimated on the basis of this period is given in the top right corner of the display.

Figure 4

EXAMPLE OF AMPLITUDE MEASUREMENT OF THE PLOSIVE BURST

FOR /d/ IN /dat/ SPOKEN BY AN ADULT MALE


Screen Files Edit Analysis Record Play Quit lo Curs

```
CH 1     0.923 Volts  Init  =  538.615    Length =    2.538  Freq =   393.9 Hz
```

Screen Files Edit Analysis Record Play Quit lo Curs

```
CH 1     3.433 Volts  Final =  560.923    Length =   22.346  Freq =    44.8 Hz
```

-- In the top figure, the plosive burst for /d/ has been isolated from the rest of the waveform. The cursor is placed on the point of greatest amplitude, as indicated by the amplitude value (in Volts) in the top left corner of the display. In the bottom figure, the plosive burst and the first pitch period of the vowel are displayed; the amplitude value is displayed for measurement of the pitch period.

The point of greatest amplitude of the vowel was visually identified by the investigator from the waveform display and confirmed by measurement with the cursor. The amplitude (in Volts) of this vowel peak was automatically calculated and displayed. The ratio of the highest amplitude peak of the plosive burst to the highest amplitude point of the waveform of the postconsonantal vowel was used in the statistical comparisons.

Perceptual Analysis

The procedure for the perceptual analysis of the speaker-subjects' word productions has been adapted from that described by Forrest and Rockman (1988) and Hodge (1989). Five speech-language pathologists served as listener subjects. They were native speakers of Canadian English with no history of speech or hearing deficits. Their hearing acuities were within normal limits as measured by means of standard audiometric screening procedures.

The perceptual validation task was administered and analyzed by computer using customized experimental software (PERCEPT) designed to operate with the CSRE (Jamieson, 1989) speech analysis program on an IBM - PC/AT microcomputer. The CSpeech files of speaker-subjects' utterances of "tot" and "dot" were converted to CSRE files and then reconstituted via a D/A converter (Data Translation 2801), low-pass filtered at 10.0 kHz (Frequency Devices 901), amplified (Realistic SA-150 integrated stereo amplifier) and presented to the listeners free field in a double-walled, sound-insulated test chamber (Amplifon, Ltd.).

Listeners were seated at a table in the sound-insulated chamber with access to the computer keyboard and a computer mouse, and facing a window into the control room. The computer monitor was positioned outside the window so that it was visible to the listener, but its fan noise was not audible. Listener subjects heard the single words at a comfortable listening level. Only one listener performed the listening task at a time.

Listeners were informed in advance as to the age group of the speakers to whom they would be listening. Presentation of a subset of 10 practice items preceded presentation of the experimental words in each block to familiarize the listeners with the voice and speech characteristics of the new age group. In this way, judgments of voicing were not influenced by changing pitch characteristics.

Each word was presented by the computer twice consecutively. The interstimulus interval between the initial presentation and the repetition was one second. Presentation of the subsequent word was delayed by the computer until the listener responded. Response options were displayed in 'boxes' on the computer screen. Listeners were asked to identify the initial plosive of each word they heard as one of four possible choices: (a) /d/ -- indicating that the listener heard a definite voiced alveolar plosive; (b) /d?/ -- indicating that the listener heard a possible voiced alveolar plosive; (c) /t?/ -- indicating that the listener heard a possible voiceless alveolar plosive; or (d) /t/ -- indicating that the listener heard a definite voiceless alveolar plosive. The ambiguous choices were included to allow the listener to indicate uncertainty while still forcing a choice. Listeners indicated their choices by moving a cursor driven by the computer mouse to the box corresponding to /d/, /d?/, /t?/, or /t/ on the computer screen. The computer automatically stored and organized the listener-subjects' performances according to presentation order and listener choice.

Words were presented in groups according to speaker age (i.e., all words produced by a given speaker age group were introduced before the words produced by another age group were presented). The 10 words spoken by each speaker subject within an age group were randomly distributed throughout each block. Therefore, listeners heard four blocks of 220 words corresponding to the four speaker groups. Order of presentation of the four blocks of words varied for each listener to avoid potential effects of order or fatigue. A total of 920 words were presented to the listeners (i.e., 800 tokens + 80 repeated tokens for intra-listener reliability + 40 practice tokens). In consideration of the high number of tokens being presented, listeners were given a 15-minute break after analyzing half of the tokens. The experimental session, including the break, lasted approximately 1 1/2 hours for each listener.

Reliability

Intra-listener reliability.

A mutually exclusive set of 80 tokens, 20 tokens from the data of each of the four age groups, was randomly selected to be presented twice in the perceptual validation task so that a measure of intra-listener reliability could be calculated (Hodge, 1989). Measures of intra-listener reliability were calculated for the number of complete agreements and for the number of within-category agreements. Complete agreements were defined as instances in which the listener's first and second judgments of the repeated word were exactly the same (e.g., first and second judgments = /d/). Within-category agreements were defined as instances in which the listener's first and second judgments were within the same voicing category (e.g., first judgment = /d/, second judgment = /d?/). Only the listener's reliability of judgment was evaluated with this procedure, not the listener's accuracy.

Acoustic measures reliability.

Intra-rater reliability for accuracy of acoustic measurement was calculated for 10% of the speaker-subjects' data. The speech samples from two speakers in each age group were randomly selected. All of the acoustic variables of interest for each word produced by these two subjects were remeasured. Measures of VOT were considered accurate if repeated measures were within one pitch period interval. The time interval equivalent to one pitch period differed according to the speaker age. Therefore, mean pitch period intervals (in ms) were calculated according to each speaker's fundamental frequency. The interval appropriate to the speaker's age was used as the criterion of accuracy for repeated measures of VOT.

Measures of plosive burst and peak vowel amplitude were considered accurate when the first and second measures were in exact agreement. This stringent criterion was applied as the calculation of amplitude (in Volts) provided by the CSpeech software program should be identical if the appropriate point in the waveform display has been selected.

Measures of F1 were considered accurate if they were within Lindblom's (1963) estimate of an expected formant frequency measurement error of one-quarter the fundamental frequency. Measures of F0 were considered accurate if differences between first and second measurements were within the frequency range indicated by one cursor increment. These frequency ranges varied depending upon the specified analyzing window and the fundamental frequency of the speaker. Estimates of the magnitude of measurement error were calculated for all acoustic variables of interest (Hodge, 1989).

Data Analysis

Perceptual Data

Originally, six listeners participated in the perceptual validation task. The results for only five listeners were intended for inclusion in the statistical analyses. The sixth listener participated to provide feedback to the principal investigator about the perceptual validation task. However, due to the experimenter's error in saving the fourth listener's data, that listener's judgments of the adult speakers' data were lost. Consequently, perceptual results from the "fourth" listener included in the statistical analyses were actually a combination of the results from two listeners (those for the fourth listener for all of the children's data, and those for the sixth listener for the adults' data). Substitution of the complete data set for the sixth listener in place of that for the fourth listener was considered to be inappropriate, as the sixth listener had significantly more experience in perceptual experiments with children's voices; therefore, her judgments of the children's data could have differed from those of the other listeners.

Intra-listener reliability.

As mentioned previously, measures of intra-listener reliability were calculated for the number of complete agreements and for the number of within-categorv agreements made by the listeners. The number of complete and within-category agreements according to age group and voicing condition was calculated for each listener. Additionally, 2-way univariate analyses of variance with repeated measures on one level were used to compare intra-listener reliability rates among the speaker-subject age groups and between the two voicing conditions. Although it is acknowledged that the employment of an intra-class correlation may be a more commonly used statistical analysis for reliability measures, the ANOVA analysis was chosen as it allowed comparisons of differences that might have existed in a listener's perception of the voicing distinction among the speaker age groups. The independent variables included (1) speaker age group with four levels (2.5 year old children, 4.5 year old children, 10 year old children and adults), and (2) voicing condition with two levels (voiced word-initial alveolar plosives and voiceless word-initial alveolar plosives). The dependent variable was either the number of complete agreements or the number of within-category agreements provided by the listeners. Separate univariate analyses of variance were conducted for each dependent variable. The per-experiment alpha level was set at .05, and the

per-comparison alpha level was adjusted to .01 to account for multiple univariate comparisons (Kirk, 1982).

As mentioned previously, the data for the fourth listener were actually a combination of the results from two listeners. The decision to combine the data for inclusion in these univariate analyses of variance was deemed to be acceptable given the assumptions that (a) all the listeners were trained and knowledgeable and (b) listeners' judgments of the adults' data should not have been as variable as those of the children's data.

Inter-listener comparisons.

Descriptive statistical data were tabulated describing the mean number and range of values of correctly perceived, incorrectly perceived, and ambiguous tokens according to age group and voicing category. Correct listener judgments were defined as each occurrence in which the word-initial alveolar plosive was identified by a listener as it was intended by the speaker with respect to the voicing feature (e.g., the speaker's production of /t/ was identified by the listener as /t/). Incorrect judgments were defined as each occurrence in which the word-initial alveolar plosive was identified by a listener as opposite to the way the speaker intended with respect to the voicing feature (e.g., the speaker's production of /t/ was identified by the listener as /d/). Ambiguous judgments were defined as each occurrence in which the word-initial alveolar plosive was identified by the listener as a questionable /t/ or questionable /d/.

Univariate analyses of variance with repeated measures were used to compare the listeners' judgments of the word initial alveolar plosives across speaker age groups and between voicing conditions. All acceptable productions of the experimental words produced by each speaker-subject were included in the statistical analyses. Only 2 of 800 words were excluded; both of these were from the data set of the youngest age group. In one case, the child did not produce a fifth repetition of "dot", and in the other case the word produced was incorrect (i.e., the child said "dog" instead of "dot").

Initially, three 3-way univariate analyses of variance were conducted, one for each dependent variable. Independent variables included (1) gender with two levels (male and female), (2) speaker age group with four levels (2.5 year old children, 4.5 year old children, 10 year old children, and adults), and (3) voicing condition with two levels (voiced word-initial alveolar plosives and voiceless word-initial alveolar plosives). The dependent variable was one of the following: the mean number of correct, incorrect or ambiguous judgments provided by the listeners. The per-experiment alpha level was set at .05, and the per-comparison alpha level was adjusted to .01 to account for multiple univariate comparisons (Kirk, 1982). As no significant gender differences were found for the perceptual analysis, this variable was omitted from subsequent statistical analyses. Therefore, three 2-way univariate analyses of variance were conducted. When main effects were significant, Scheffe's Test of comparisons was use to determine where the significant differences occurred. This posthoc test of comparisons is considered to be more stringent than others, such as the Tukey test, thus reducing the probability of Type 1 errors (Bruning & Kintz, 1987).

Based on visual inspection of the data for each listener, it appeared that the fourth listener's judgments of the children's data differed from those of the other listeners. Consequently, each of the previously mentioned univariate analyses was repeated excluding the data from the fourth listener. Results from both sets of analyses are reported in the Results section.

Acoustical Data

Descriptive statistical data were tabulated describing the mean values and standard deviations for each acoustic variable of interest according to age group and voicing condition. In the case of the spectral variables (i.e., F1 and F0) such data were obtained for frequency estimates in Hz, and Bark and log transformed values.

In order that statistical comparisons could be made among the various speaker age groups, values obtained from only one of the two spectral normalization transformations could be included in subsequent statistical analyses. Using the frequency data first in Hertz, then in Bark, and finally in natural log as the dependent measure, the experimenter performed three separate 2-way univariate analyses of variance with repeated measures to decide which forms of the frequency data to use in subsequent analyses. It was hypothesized that differences in frequency in the low frequency ranges could be larger using the Bark transformation than could be expected to occur using a logarithmic transformation (Miller, 1989; Nearey, 1989). If this occurred, any statistically significant differences among spectral values transformed to the Bark scale could be artificial. As the results from the statistical analyses were comparable, only values transformed to the log scale were included in subsequent statistical analyses.

Experimental questions 2 through 5, which considered differences that may exist in the various acoustic variables relative to voicing condition and speaker age, were answered within a mixed 2-factor (4 x 2) multivariate design with repeated measures on the factor of voicing condition. The independent variables included (1) speaker age group with four levels and (2) voicing condition with two levels. The four dependent variables included VOT, F1 at onset of the postconsonantal vowel, F0 at onset of the postconsonantal vowel, and plosive burst amplitude relative to the amplitude of the postconsonantal vowel. The MANOVA analysis was appropriate as multiple independent and dependent variables were involved. It was advantageous to the execution of multiple univariate analyses of variance because the latter would result in higher probabilities of Type I errors than indicated by the level of significance used.

One perceptually-validated token for each plosive (/t/ or /d/) produced by each speaker-subject was selected for inclusion in the MANOVA. Consequently, equal numbers of data points across age groups were assured. A perceptually validated token was defined as any token that was correctly identified by at least four out of five listeners. The selection process for the tokens occurred in the following order: First choice for selection was any token correctlv identified by 5/5 listeners; if such a token was not available, a token correctly identified by 4/5 listeners was selected; and if multiple tokens that met the selection criteria existed for any speaker, one token was randomly selected. In all but one case, perceptually-validated tokens existed for inclusion in the analysis. For one of the speakers in the youngest age group, there were no perceptually-validated tokens for his production of "dot". Consequently, the token selected for inclusion in the statistical analysis was correctly perceived by 3/5 listeners and within-category for the remaining two listeners.

The MANOVA analysis yielded information about two main effects and one first-order interaction between speaker age and voicing condition. Main effects that were significant at the .05 level were subsequently analyzed via multiple univariate analyses of variance and appropriate posthoc comparisons were conducted to identify the sources of significant effects. For these univariate analyses, the per-experiment alpha level was set at .05, and the per comparison alpha level was adjusted to .01 to account for multiple comparisons.

The remaining experimental question, which explored the relative contribution of the acoustic cues associated with the voicing contrast to listeners' perceptions of word initial alveolar plosives, was addressed through a descriptive correlational design using a simple discriminant function analysis. This analysis generated one discriminant function equation to allow comparisons of the levels of independent (predictor) variables relative to each dependent (criterion) variable (Huck, Cormier, & Bounds, 1974). Separate discriminant function analyses could not be conducted for each age group due to insufficient numbers of tokens. Specifically, as one of the assumptions of the discriminant function analysis is that the data samples are independent, only one word from each subject could be included in the analysis (i.e., either 'dot' or 'tot'); consequently, only 20 words from any one age group could be used, an insufficient number for separate discriminant function analyses with 3 predictor variables. Therefore, a single discriminant function analysis was conducted with one perceptually-validated token from each of 80 subjects included in the analysis. Equal numbers of 'dot' and 'tot' were represented (i.e., 40 productions of 'dot' and 40 productions of 'tot').

Three acoustic variables served as the predictor variables: VOT, plosive burst amplitude, and the difference between F1 and F0. The criterion variable, plosive voicing status, had two levels: identification of the word-initial plosive as either /t/ or /d/. The fact that three and not four predictor variables were used deserves a comment. The separate (acoustic) variables of F1 and F0 were combined into one indepen .nt variable expressed as the di ference between F1 and F0 for this analysis. The F1 - F0 difference provided information on the combined contribution of spectral measures to the perception of voicing This was a necessary adjustment to the acoustical variables because the conversion of spectral values expressed in Hertz to either the Bark or log scales was not sufficient to normalize the frequency values for vocal tract size differences. In either scale, spectral values for the speech of younger speakers were of higher frequency than those for older speakers. A complete normalization of values resulted from expressing the transformed values as a difference of one value from another.

A simple discriminant function analysis was chosen because the criterion variable, categorization of the word initial plosive as either /t/ or /d/, was a nominal data form. The analysis generated one discriminant function prediction equation, with coefficients corresponding to each predictor variable. The Wilk's lamda was converted to a chi-square distribution to allow a test of significance to determine whether the discriminant function prediction equation facilitated more accurate prediction of listeners' judgments of voicing than would be possible by chance alone (Huck et al., 1974). The computerized statistical analysis package SPSS-X automatically calculated standardized discriminant function values which allowed interpretation of the relative importance of each predictor variable to separation of the criterion variables (McLaughlin, 1980). Finally, prior to the classification phase of the analysis, the Box's M Test was employed to determine homogeneity of covariance. Estimates of the percentage of correct classifications were calculated for the 80 tokens used to derive the discriminant function prediction equation. Additionally, a cross-validation sample of 80 tokens not included in the original analysis, was used to test the stability of the classification scheme; the percentage of correct classifications for this hold-out sample also was calculated.

Results

Perceptual Data

Intra-Listener Reliability

The number of complete agreement (in which the listener's first and second judgments of a repeated word were exactly the same) and within-category agreements (in which the listener's first and second judgments were within the same voicing category) were calculated for each listener according to speaker age group and voicing condition. Tables 1 and 2 summarize the percentage of complete and within-category agreements, respectively, for each listener. Additionally, the ratio of actual agreements to total possible agreements is provided. As mentioned previously, a total of 80 tokens was repeated in the perceptual task, 20 from each age group, to allow comparisons of intra-listener reliability. Of these 80 tokens, equal numbers of voiced and voiceless tokens were represented. However, the proportion of voiced and voiceless tokens *within* an age group was not necessarily equal; for example, 14 voiced and 6 voiceless tokens from the data set of the 2.5 year old speakers were randomlv selected for repetition by the CSRE computer program while equal numbers of voiced and voiceless tokens were randomly selected from the data set of the 4.5 year old speakers.

Visual inspection of these Tables indicated that listeners' reliability scores were higher for within-

Table 1

Intra-rater reliability scores (%) for complete agreements for each listener across speaker age and by voicing condition. Ratio of accurate tokens to possible tokens is given in parenthesis.

	Speaker Age-Level				
Listener	**2.5 Years**	**4.5 Years**	**10 Years**	**Adult**	**Grand Mean**
L1	80% (16/20)	95% (19/20)	95% (19/20)	100% (20/20)	93% (74/80)
L2	90% (18/20)	95% (19/20)	100% (20/20)	100% (20/20)	96% (77/80)
L3	95% (19/20)	85% (17/20)	95% (19/20)	85% (17/20)	90% (72/80)
L4[1]	60% (12/20)	85% (17/20)	75% (15/20)	100% (20/20)	80% (64/80)
L5	95% (19/20)	90% (18/20)	100% (20/20)	100% (20/20)	96% (77/80)
Grand Mean by Speaker Age Group	84% (84/100)	90% (90/100)	93% (93/100)	97% (97/100)	91% (364/400)

[1] Listener 4 data are a combination of those from two listeners; one judged the children's data and the other judged the adults' data.

Table 2

Intra-rater reliability scores (%) for within-category agreements for each listener across speaker age and by voicing condition. Ratio of accurate tokens to possible tokens is given in parentheses.

	Speaker Age-Level				
Listener	**2.5 Years**	**4.5 Years**	**10 Years**	**Adult**	**Grand Mean**
L1	95% (19/20)	100% (20/20)	95% (19/20)	100% (20/20)	98% (78/80)
L2	100% (20/20)	95% (19/20)	100% (20/20)	100% (20/20)	99% (79/80)
L3	100% (20/20)	95% (19/20)	100% (20/20)	100% (20/20)	99% (79/80)
L4[1]	85% (17/20)	95% (19/20)	100% (20/20)	100% (20/20)	95% (76/80)
L5	100% (20/20)	100% (20/20)	100% (20/20)	100% (20/20)	100% (80/80)
Grand Mean by Speaker Age Group	96% (96/100)	97% (97/100)	99% (99/100)	100% (100/100)	98% (392/400)

[1] Listener 4 data are a combination of those from two listeners; one judged the children's data and the other judged the adults' data.

category agreements than for complete agreements, particularly for the youngest speaker group. This was anticipated because the complete agreement category applied a more stringent criterion. Data for complete agreements indicated that listeners' reliability scores generally improved as speaker age increased. This effect was markedly reduced in the data for within-category agreements in which reliability scores for all age groups were high, ranging from 85% to 100%.

There were some differences in the reliability data among the individual listeners. As mentioned previously, perceptual results from the "fourth" listener included in the statistical analyses were actually a combination of the results from two listeners (one who judged all of the children's data and one who judged the adults' data). Reliability results for L4 (adults) were comparable to those for the other listeners for the adult data; however, reliability results for L4 (children) were noticeably different. Reliability scores of complete agreements for L4 (children) were lower than these scores for other listeners, particularly for judgments of /d/. L4 (children), and L3 to a lesser extent, made fewer complete agreements for /d/ compared to /t/ when judging the word productions of the children. Reliability scores for within-category agreements for L4 (children) were comparable to those for other listeners, suggesting that she used the [illegible] categories

('t?' and 'd?') more frequently than did the other listeners.

Mean intra-listener reliability scores of complete agreements for each speaker age group and both voicing conditions are reported in Table 3. The mean reliability score of 5.00 is based on the total possible number of complete agreements for each token by five listeners, averaged across the 20 tokens used in the reliability measurement for each age group. A 2-way univariate analysis of variance was used to compare intra-listener reliability rates among the speaker-subject age groups and between the two voicing conditions. Differences among the reliability scores for the four speaker age groups approached but did not achieve significance (*F* [3, 72] = 2.35, *p* =.08). Similarly, differences in the reliability scores between voicing conditions approached but did not achieve significance (F [1, 72] = 3.65, *p* = .06). The interaction of speaker age and voicing condition was not statistically significant.

Mean intra-listener reliability scores of within-category agreements for each speaker age group and both voicing conditions are reported in Table 4. The mean reliability score of 5.00 for within-category agreements was calculated in the same manner as previously described for complete agreements. A 2-way univariate analysis of variance did not reveal statistical significance for either

Table 3

Mean scores and standard deviations of complete agreements among listeners across speaker age and between voicing condition; N = 5 Listeners; total possible mean reliability score is 5.000[1]. Percentage of complete agreements is given in parentheses.

		Speaker Age-Level			
Voicing		2.5 Years	4.5 Years	10 Years	Adult
/d/	M	4.214 (84%)	4.300 (86%)	4.375 (88%)	4.875 (98%)
	S.D	0.802	1.059	0.744	0.354
/t/	M	4.167 (83%)	4.700 (94%)	4.833 (97%)	4.833 (97%)
	S.D	1.169	0.483	0.389	0.389

[1] The mean reliability score of 5.00 is based on the total possible number of complete agreements for each token by 5 listeners, averaged across 20 tokens per age group.

Table 4

Mean scores and standard deviations of within-category agreements among listeners across speaker age and between voicing condition; N = 5 Listeners; total possible mean reliability score is 5.000[1]. Percentage of within category agreements is given in parentheses.

		Speaker Age-Level			
Voicing		2.5 Years	4.5 Years	10 Years	Adult
/d/	M	4.857 (97%)	4.800 (96%)	4.875 (98%)	5.000 (100%)
	S.D	0.363	0.422	0.354	0.000
/t/	M	4.667 (93%)	4.900 (98%)	5.000 (100%)	5.000 (100%)
	S.D	0.817	0.316	0.000	0.000

[1] The mean reliability score of 5.00 is based on the total possible number of within-category agreements for each token by 5 listeners, averaged across 20 tokens per age group.

main effects of speaker age or voicing condition, or the interaction of speaker age and voicing condition.

Inter-listener Comparisons

Descriptive data on inter-listener comparisons were compiled to include the mean number, standard deviation, and percentage of correctly perceived, incorrectly perceived, and ambiguous tokens according to age group and voicing category. Tables 5, 6, and 7 summarize individual listeners' mean scores for correct, incorrect, and ambiguous judgments, respectively. Mean scores are reported out of a total possible 5.000. This value represents the number of correct, incorrect, or ambiguous judgments out of 5 repetitions of 'dot' or 'tot' produced by each speaker and averaged across the 20 speakers in each age group.

A number of trends are observable among these data. Firstly, regarding correct judgments, listeners made the least number of correct judgments for sounds produced by the youngest speaker group. Additionally, most listeners made fewer correct judgments of children's productions of /d/ than /t/ whereas they made more correct judgments of adults' productions of /d/ than /t/. However, if the data from L4 (children) are eliminated, this tendency is not as evident. L4 (children) made markedly fewer correct judgments of /d/ than /t/ as compared to the correct judgments of the other listeners.

Table 5

Mean scores and standard deviations of listeners' correct judgments across speaker age and between voicing conditions. Total possible mean score of correct judgments is 5.00[1]. Percentage of correct judgments is given in parentheses.

		Speaker Age-Level							
Listener		2.5 Years		4.5 Years		10 Years		Adult	
		d	t	d	t	d	t	d	t
L1	M.	3.60 (72%)	3.60 (72%)	4.85 (97%)	4.75 (95%)	4.95 (99%)	4.95 (99%)	5.00 (100%)	4.95 (99%)
	S.D.	1.47	1.35	0.37	0.91	0.22	0.22	0.00	0.22
L2	M.	3.80 (76%)	4.25 (85%)	4.40 (88%)	4.85 (97%)	4.65 (93%)	5.00 (100%)	5.00 (100%)	4.90 (98%)
	S.D.	1.11	1.07	0.82	0.37	0.59	0.00	0.00	0.45
L3	M.	4.00 (80%)	3.45 (69%)	4.60 (92%)	4.45 (89%)	4.55 (91%)	4.35 (87%)	4.55 (91%)	3.80 (76%)
	S.D.	1.08	1.15	0.82	0.95	0.95	1.04	0.61	1.61
L4[2]	M.	2.10 (42%)	4.00 (80%)	3.60 (72%)	4.70 (94%)	2.70 (54%)	3.95 (79%)	5.00 (100%)	4.85 (97%)
	S.D.	1.37	1.30	1.54	0.92	1.08	1.23	0.00	0.49
L5	M.	4.10 (82%)	4.25 (85%)	4.75 (95%)	4.90 (98%)	4.55 (91%)	5.00 (100%)	5.00 (100%)	4.95 (99%)
	S.D.	1.25	1.16	0.55	0.45	1.00	0.00	0.00	0.22

[1] The mean score of 5.00 for each listener is based on the total possible number of correct judgments across 5 repetitions of "tot" or "dot" produced by each speaker, averaged over 20 speakers.

[2] Listener 4 data are a combination of those from two listeners; one judged the children's data and the other judged the adults' data.

Table 6

Mean scores and standard deviations of listeners' incorrect judgments across speaker age and between voicing conditions. Total possible mean score of incorrect judgments is 5.00[1]. Percentage of incorrect judgments is given in parentheses.

		Speaker Age-Level							
Listener		2.5 Years		4.5 Years		10 Years		Adult	
		d	t	d	t	d	t	d	t
L1	M.	0.50 (10%)	0.20 (4%)	0.05 (1%)	0.05 (1%)	0.05 (1%)	0.00 (0%)	0.00 (0%)	0.00 (0%)
	S.D.	1.05	0.52	0.22	0.22	0.22	0.00	0.00	0.00
L2	M.	0.40 (8%)	0.15 (3%)	0.05 (1%)	0.00 (0%)	0.00 (0%)	0.00 (0%)	0.00 (0%)	0.00 (0%)
	S.D.	0.88	0.49	0.22	0.00	0.00	0.00	0.00	0.00
L3	M.	0.35 (7%)	0.20 (4%)	0.00 (0%)	0.00 (0%)	0.00 (0%)	0.00 (0%)	0.00 (0%)	0.00 (0%)
	S.D.	0.67	0.52	0.00	0.00	0.00	0.00	0.00	0.00
L4[2]	M.	0.45 (9%)	0.05 (1%)	0.05 (1%)	0.00 (0%)	0.00 (0%)	0.00 (0%)	0.00 (0%)	0.00 (0%)
	S.D.	0.76	0.22	0.22	0.00	0.00	0.00	0.00	0.00
L5	M.	0.30 (6%)	0.30 (6%)	0.05 (1%)	0.00 (0%)	0.00 (0%)	0.00 (0%)	0.00 (0%)	0.00 (0%)
	S.D.	0.66	0.66	0.22	0.00	0.00	0.00	0.00	0.00

[1] The mean score of 5.00 for each listener is based on the total possible number of incorrect judgments across 5 repetitions of "tot" or "dot" produced by each speaker, averaged over 20 speakers.

[2] Listener 4 data are a combination of those from two listeners; one judged the children's data and the other judged the adults' data.

Table 7

Mean scores and standard deviations of listeners' ambiguous judgments across speaker age and between voicing conditions. Total possible mean score of ambiguous judgments is 5.00[1]. Percentage of ambiguous judgments is given in parentheses.

		Speaker Age-Level							
Listener		2.5 Years		4.5 Years		10 Years		Adult	
		d	t	d	t	d	t	d	t
L1	M.	0.85 (17%)	1.15 (23%)	0.10 (2%)	0.20 (4%)	0.00 (0%)	0.05 (1%)	0.00 (0%)	0.05 (1%)
	S.D.	0.99	1.18	0.31	0.70	0.00	0.22	0.00	0.22
L2	M.	0.75 (15%)	0.55 (11%)	0.55 (11%)	0.15 (3%)	0.35 (7%)	0.00 (0%)	0.00 (0%)	0.10 (2%)
	S.D.	0.97	0.76	0.83	0.37	0.59	0.00	0.00	0.45
L3	M.	0.60 (12%)	1.30 (26%)	0.40 (8%)	0.55 (11%)	0.45 (9%)	0.65 (13%)	0.45 (9%)	1.20 (24%)
	S.D.	0.75	0.92	0.82	0.95	0.95	1.04	0.61	1.61
L4[2]	M.	2.40 (48%)	0.90 (18%)	1.35 (27%)	0.30 (6%)	2.30 (46%)	1.05 (21%)	0.00 (0%)	0.15 (3%)
	S.D.	1.14	1.29	1.50	0.92	1.08	1.23	0.00	0.49
L5	M.	0.55 (11%)	0.40 (8%)	0.20 (4%)	0.10 (2%)	0.45 (9%)	0.00 (0%)	0.00 (0%)	0.05 (1%)
	S.D.	0.89	0.68	0.52	0.45	1.00	0.00	0.00	0.22

[1] The mean score of 5.00 for each listener is based on the total possible number of ambiguous judgments across 5 repetitions of "tot" or "dot" produced by each speaker, averaged over 20 speakers.

[2] Listener 4 data are a combination of those from two listeners; one judged the children's data and the other judged the adults' data.

Secondly, regarding incorrect judgments, listeners made few incorrect judgments for any age group. However, as hypothesized, the number of incorrect judgments decreased as speaker age increased. The largest number of incorrect judgments were made for sounds produced by the youngest speakers. Additionally, listeners made more incorrect judgments of children's productions of /d/ than /t/, particularly for the youngest speakers.

The response class data for ambiguous judgments were particularly affected by the responses for L4 (children). All listeners made more ambiguous judgments for sounds produced by 2.5 year old speakers. If all five listeners were included, there were more ambiguous judgments evident for children's productions of /d/ than /t/, and fewer ambiguous judgments for adults' productions of /d/ than /t/. If the data from the fourth listener were not included, however, the pattern of ambiguous judgments was similar for the 2.5 year old and adult speakers. Specifically, there were fewer ambiguous judgments of /d/ than /t/ for both 2.5 year old and adult speakers. There continued to be more ambiguous judgments of /d/ than /t/ produced by 4.5 year old and 10 year old speakers. Overall, if the data from L4 were eliminated, there were fewer ambiguous judgments compared to the data from 5 listeners (i.e., the range of ambiguous judgments by L4 [children] across speaker age was 42% compared to a maximum range of 23% for any other listener).

Three 2-factor (4 x 2) univariate analyses of variance with repeated measures on the factor of voicing condition were used to compare the listeners' judgments of the word-initial alveolar plosives across speaker age groups and between voicing conditions. All acceptable productions of the experimental words produced by each speaker-subject were included in the statistical analyses. As described in the previous paragraphs, there were differences in the patterns of perceptual judgments between L4 (children) and the other listeners. In order to test whether L4 (children) had an inordinate effect upon the results, statistical analyses were conducted with data from all five listeners and repeated excluding the data from the fourth listener.

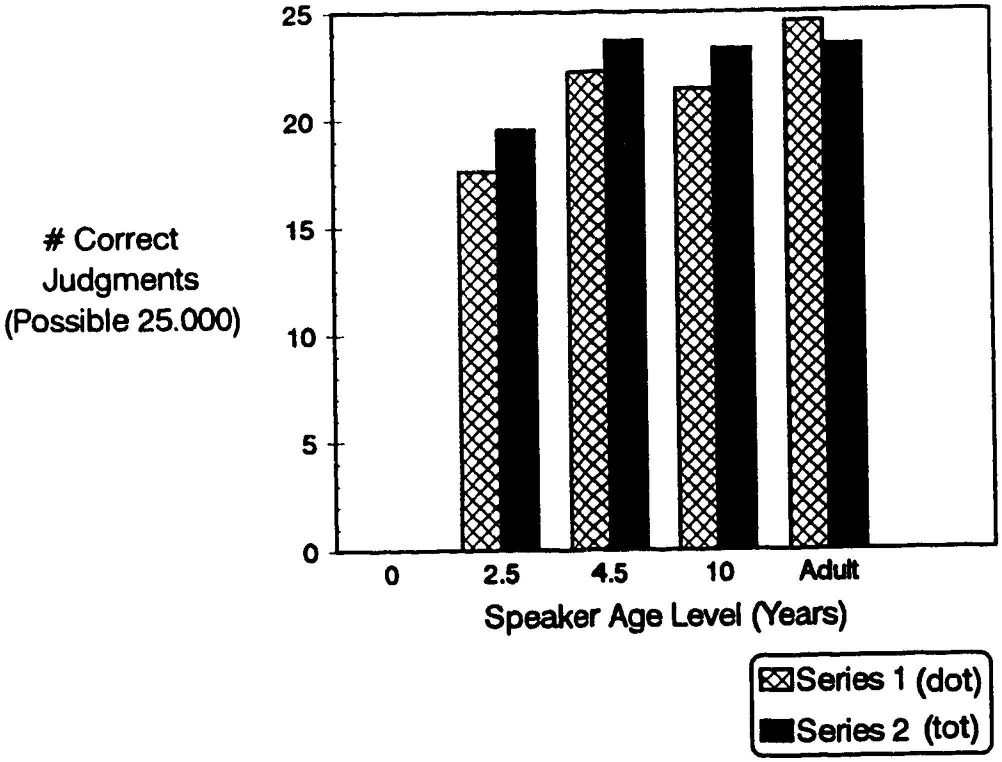
Figure 5 graphs the results from the univariate analysis of correct judgments made by all five listeners. The mean and standard deviations are also provided for each age group and voicing condition. The mean score of 25.00 given in Figures 5 - 7 is based on the total possible number of correct, incorrect, or ambiguous judgments made by 5 listeners across 5 repetitions of 'dot' or 'tot' produced by each speaker, averaged across 20 speakers per speaker age group. There were significant differences in the number of correct judgments among speaker age group (F [3, 76] = 13.71, p = .0003). Listeners made fewer correct judgments of sounds produced by 2.5 year old speakers compared to sounds produced by 4.5 year old, 10 year old and adult

Figure 5

Mean scores and standard deviations of correct judgments of /t/ and /d/ for five listeners across speaker age and between voicing conditions. Total possible mean score is 25.000[1].

Speaker Age-Level

Voicing		2.5 Years	4.5 Years	10 Years	Adult
/d/	M. S.D.	17.600 (5.295)	22.200 (2.840)	21.400 (3.085)	[illegible] (0.605)
/t/	M. S.D.	19.550 (5.165)	23.650 (3.345)	23.250 (1.970)	23.450 (2.417)



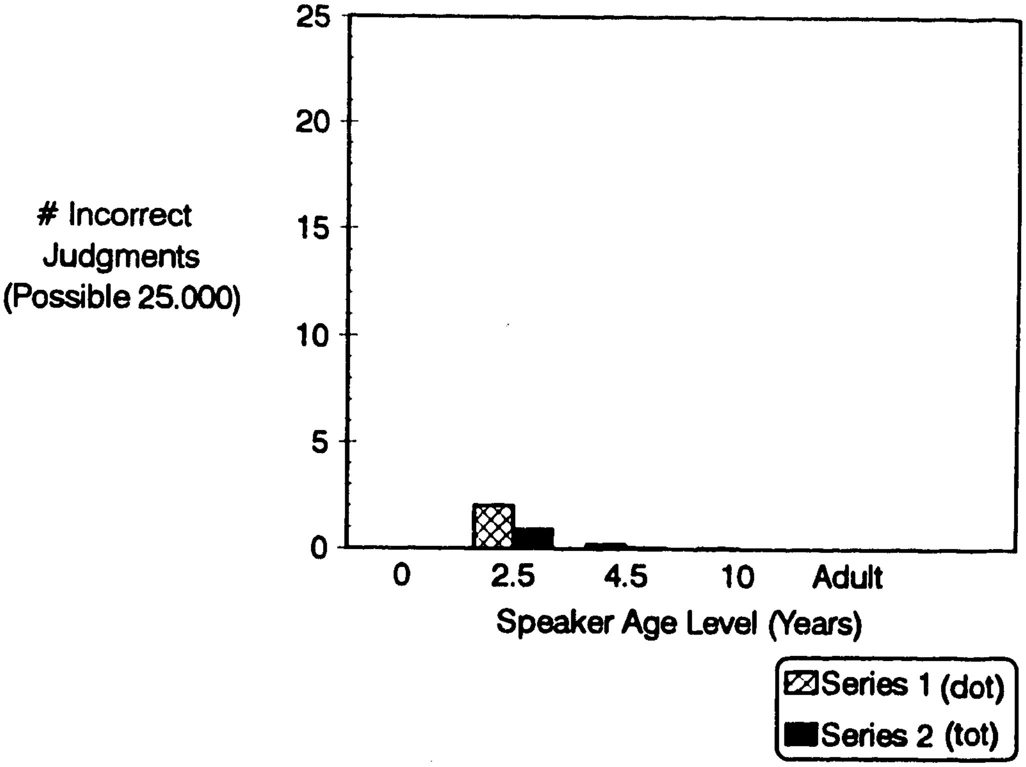
[1] The mean score of 25.000 is based on the total possible number of correct judgments by 5 listeners across 5 repetitions of "tot" or "dot" produced by each speaker, averaged over 20 speakers.

Figure 6

Mean scores and standard deviations of incorrect judgments of /t/ and /d/ for five listeners across speaker age and between voicing conditions. Total possible mean score is 25.000[1].

Speaker Age-Level

Voicing		2.5 Years	4.5 Years	10 Years	Adult
/d/	M.	2.000	0.200	0.050	0.000
	S.D.	(3.685)	(0.894)	(0.224)	(0.000)
/t/	M.	0.900	0.050	0.000	0.000
	S.D.	(2.024)	(0.224)	(0.000)	(0.000)



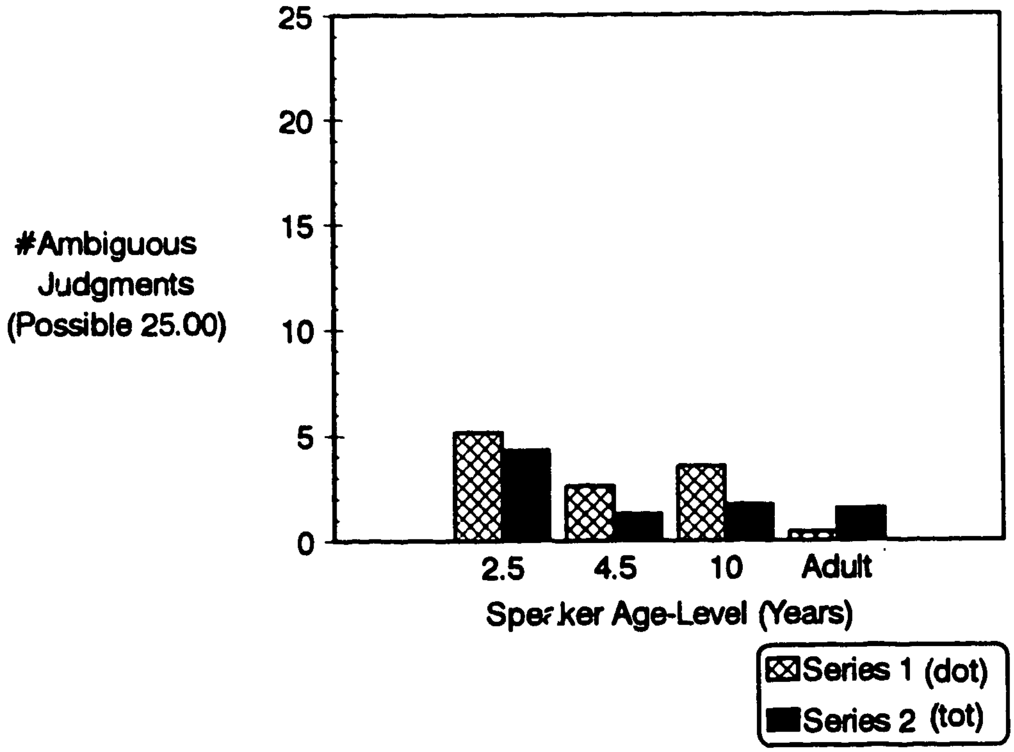
[1] The mean score of 25.000 is based on the total possible number of incorrect judgments by 5 listeners across 5 repetitions of "tot" or "dot" produced by each speaker, averaged over 20 speakers.

Figure 7

Mean scores and standard deviations of ambiguous judgments of /t/ and /d/ for five listeners across speaker age and between voicing conditions. Total possible mean score is 25.000[1].

Speaker Age-Level

Voicing		2.5 Years	4.5 Years	10 Years	Adult
/d/	M.	5.150	2.600	3.550	0.450
	S.D.	(2.852)	(2.583)	(3.086)	(0.605)
/t/	M.	4.300	1.300	1.750	1.550
	S.D.	(3.541)	(3.131)	(1.970)	(2.412)



[1] The mean score of 25.000 is based on the total possible number of ambiguous judgments by 5 listeners across 5 repetitions of "tot" or "dot" produced by each speaker, averaged over 20 speakers.

speakers. Differences in the number of correct judgments between voicing conditions approached but did not achieve significance (p = .02). Similarly, the interaction of speaker age and voicing condition approached but did not achieve significance (p = .04).

Figure 6 displays the results from the univariate analysis of incorrect judgments. The main effect of speaker age was also statistically significant in this analysis (F [3, 76] = 6.42, p = .0006). Listeners made significantly more incorrect judgments of sounds produced by 2.5 year old speakers than of sounds produced by 4.5 year old, 10 year old or adult speakers.

Finally, Figure 7 displays the results from the univariate analysis of ambiguous judgments. Again, only the main effect of speaker age was statistically significant (F [3, 76] = 11.25, p = .003). Listeners made more ambiguous judgments of sounds produced by 2.5 year old speakers tnan of sounds produced by 4.5 year old, 10 year old and adult speakers. The main effect of voicing condition (p = .05) and the interaction of speaker age and voicing condition (p = .03) approached but did not achieve significance.

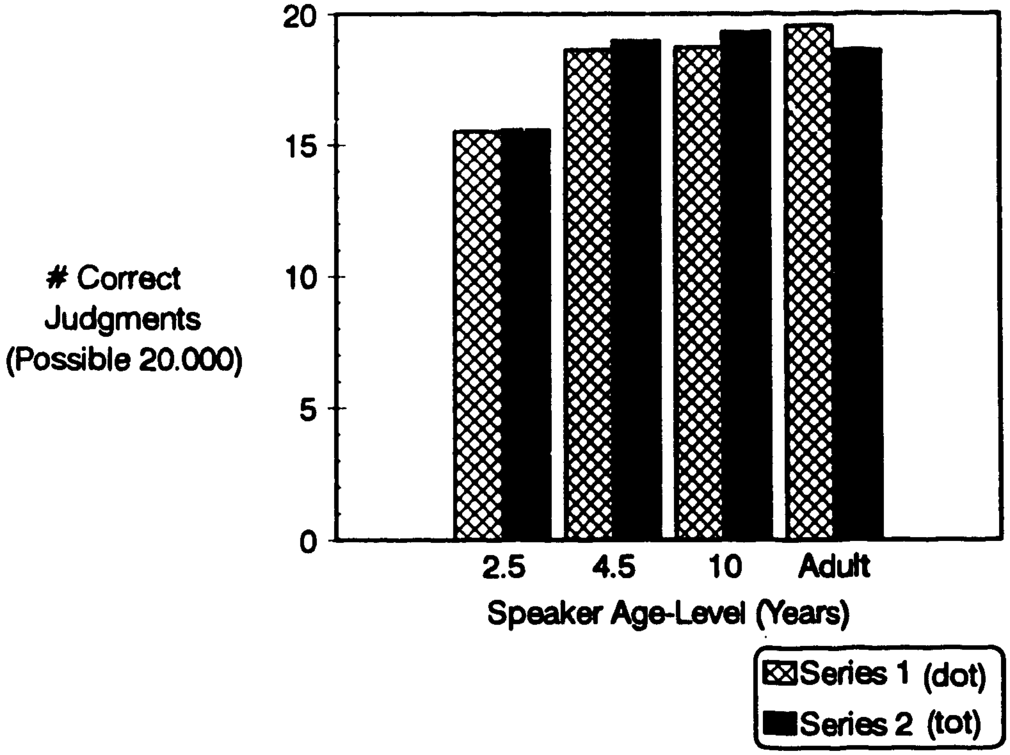
Figures 8, 9, and 10 summarize the results from the univariate analyses of correct, incorrect, and ambiguous judgments without data from L4 (children and adults). The mean score of 20.00 displayed in these figures is based on

Figure 8

Mean scores and standard de\iations of correct judgments of /t/ and /d/ for four listeners across speaker age and between voicing conditions. Total possible mean score is 20.000[1].

Speaker Age-Level

Voicing	2.5 Years	4.5 Years	10 Years	Adult
/d/	M. 15.500	18.600	18.700	19.550
	S.D. (4.335)	(1.957)	(2.297)	(0.605)
/t/	M. 15.550	18.950	19.300	18.600
	S.D. (4.059)	(2.438)	(1.081)	(2.011)



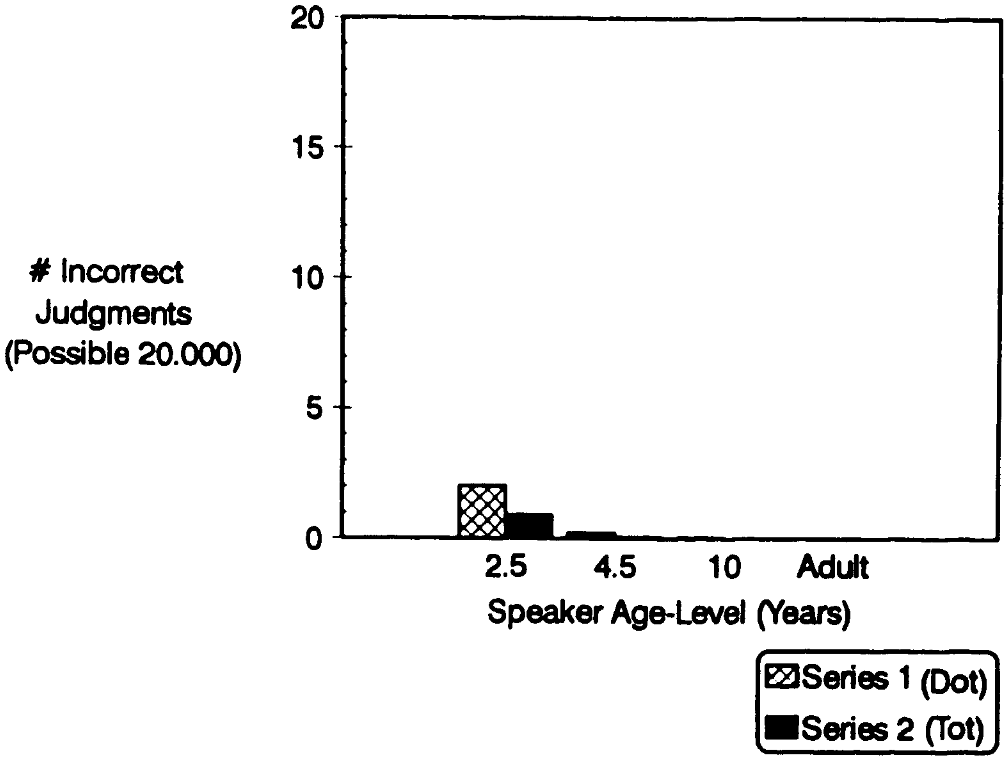
[1] The mean score of 20.000 is based on the total possible number of correct judgments by 4 listeners across 5 repetitions of "tot" or "dot" produced by each speaker, averaged over 20 speakers.

Figure 9

Mean scores and standard deviations of incorrect judgments of /t/ and /d/ for four listeners across speaker age and between voicing conditions. Total possible mean score is 20.000[1].

Speaker Age-Level

Voicing	2.5 Years	4.5 Years	10 Years	Adult
/d/	M. 1.550	0.150	0.050	0.000
	S.D. (3.017)	(0.671)	(0.224)	(0.000)
/t/	M. 0.850	0.050	0.000	0.000
	S.D. (1.927)	(0.224)	(0.000)	(0.000)



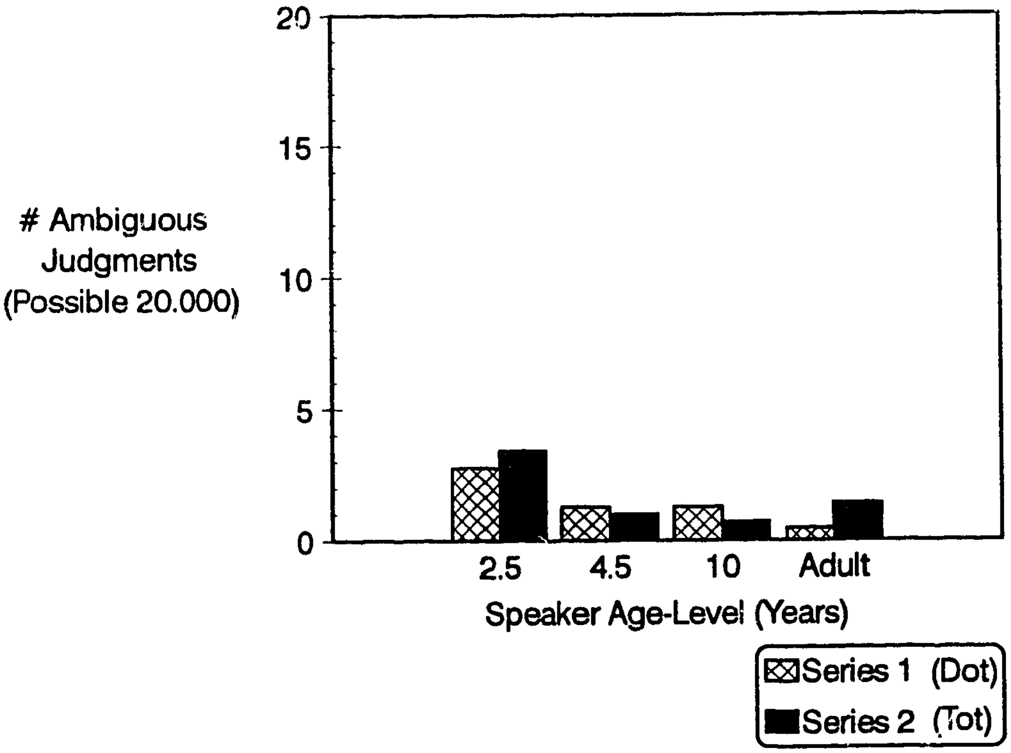
[1] The mean score of 20.000 is based on the total possible number of incorrect judgments by 4 listeners across 5 repetitions of "tot" or "dot" produced by each speaker, averaged over 20 speakers.

Figure10

Mean scores and standard deviations of ambiguous judgments of /t/ and /d/ for four listeners across speaker age and betw..en voicing conditions. Total possible mean score is 20.000[1].

Speaker Age-Level

Voicing		2.5 Years	4.5 Years	10 Years	Adult
/d/	M.	2.750	1.250	1.250	0.450
	S.D.	(2.490)	(1.773)	(2.291)	(0.605)
/t/	M.	3.400	1.000	0.700	1.400
	S.D.	(2.501)	(2.224)	(1.081)	(2.011)



[1] The mean score of 20.000 is based on the total possible number of ambiguous judgments by 4 listeners across 5 repetitions of "tot" or "dot" produced by each speaker, averaged over 20 speakers.

the total possible number of correct, incorrect, or ambiguous judgments made by 4 listeners across 5 repetitions of 'dot' or 'tot' produced by each speaker, averaged across 20 speakers per speaker age group. Generally, the results are similar to the previous analyses for listeners' correct and incorrect judgments. Regarding correct judgments, listeners made fewer correct judgments of sounds produced by 2.5 year old speakers compared to other speaker age groups (F [3, 76] = 12.08, p = .002). Listeners also made more incorrect judgments for sounds produced by 2.5 year old speakers than for sounds produced by any other age group (F [3, 76] = 6.01, p < .001). Finally, the results of the univariate analysis of ambiguous judgments indicated that listeners made more ambiguous judgments for sounds produced by 2.5 year old speakers compared to sounds produced by any other age group (F [3,76] = 8.96, p = .00004). A main effect for voicing condition and an interaction effect of speaker age and voicing condition were not statistically significant, whereas these effects approached significance when data from all five listeners were included. Therefore, it appeared that the high number of ambiguous judgments of /d/ by L4 (children), particularly for productions by the youngest speakers, had affected the data.

As mentioned previously, only perceptually-validated target words (i.e., words in which the first sound was correctly identified as 't' or 'd' by 4 of 5 listeners) were included in subsequent statistical analyses. The percentages of perceptually-validated tokens available for inclusion in these analyses varied dependent upon speaker age, with fewer tokens available for the youngest speakers compared to any other speaker group. Specifically, the percentages of perceptually-validated tokens available for inclusion in the MANOVA and discriminant function analyses were as follows (total possible number of tokens = 100):

	2.5 years	4.5 years	10 years	Adults
/d/	67%	90%	86%	100%
/t/	74%	94%	95%	96%

Acoustic Data

Acoustic Measures Reliability

Intra-rater reliability for accuracy of acoustic measurement was calculated for 10% of the speaker-subjects' data. As discussed previously, measures of VOT were considered accurate if repeated measures were within one pitch period interval. Measures of plosive burst and peak vowel amplitude were considered accurate when the first and second measures were identical. Measures of F1 were considered accurate if they varied by a value of no more than one-quarter the fundamental frequency of the subject under measurement. Measures of F0 were considered accurate if differences between first and second measurements were within the frequency range indicated by one cursor increment.

Intra-rater reliability scores were high for all acoustic variables of interest. Percentages of accurate measures of VOT ranged from 95% (2.5 year old speakers) to 100% (10 year old and adult speakers). Mean absolute measurement errors of VOT values obtained from the speech of the youngest to oldest speakers were: 0.61, 0.73, 0.41, and 0.55 ms, respectively. Percentages of accurate measures of burst amplitude ranged from 97.5% (4.5 year old and adult speakers) to 100% (2.5 and 10 year old speakers). Mean absolute measurement errors of amplitude values obtained from the speech of the youngest to oldest speakers were: 0.00, 0.02, 0.00, and 0.002 Volts, respectively. Measures of the spectral values were similarly high. Percentages of accurate measures of F0 ranged from 95% (2.5 year old speakers) to 100% (4.5 year old, 10 year old, and adult speakers). Mean absolute measurement errors of F0 values obtained from the speech of the youngest to the oldest speakers were: 2.73, 1.58, 1.05, and 0.40 Hz, respectively. Finally, percentages of accurate measures of F1 ranged from 95% (2.5 and 4.5 year old speakers) to 100% (10 year old and adult spcakers). Mean absolute measurement errors of F1 values obtained from the youngest to the oldest speakers were: 28.03, 10.15, 8.95, and 3.18 Hz, respectively. However, the ranges of absolute measurement error were quite different (i.e., 419 Hz for the 2.5 year old speakers and 13

Hz for the adult speakers).

Comparisons Among Acoustic Variables

Descriptive statistical data were tabulated describing the mean values and standard deviations according to age group and voicing condition for each acoustic variable of interest. In the case of the spectral variables (i.e., F1 and F0) data were obtained for frequency estimates in Hz, and Bark and log transformed values. Tables 8, 9, 10, and 11 summarize means and standard deviations according to speaker age and voicing condition for measures of VOT, F1, F0, and plosive burst amplitude, respectively.

The data for VOT values are summarized in Table 8. VOT values were longer for /t/ than for /d/ for all age groups. The 2.5 year old speakers displayed the greatest variability for VOT values associated with production of /t/, followed by the 4.5 year old, adult and 10 year old speakers, respectively. Adult speakers displayed the greatest variability for VOT values associated with production of /d/, followed by the 2.5 year old, 4.5 year old, and 10 year old speakers, respectively.

Table 9 summarizes means and standard deviations for measures of F1. F1 values for the postconsonantal vowel increased as speaker age decreased; this was most evident for F1 measures expressed in Hertz and least evident for F1 measures expressed as log values. F1 values for the

Table 8

Group means and standard deviations for voice onset time (in ms) across speaker age and between voicing conditions.

	Speaker Age-Level			
	2.5 Years	4.5 Years	10 Years	Adult
/d/	17.72 (15.97)	18.55 (6.80)	23.01 (7.36)	2.23 (33.46)
/t/	82.83 (33.96)	92.55 (24.56)	84.49 (14.23)	87.63 (16.10)

Table 9

Group means and standard deviations for first formant (F1) at onset of the postconsonantal vowel across speaker age and between voicing conditions.

	Speaker Age-Level			
	2.5 Years	4.5 Years	10 Years	Adult
	Values in Hertz			
/d/	899.10	761.27	651.63	564.43
	(172.10)	(67.48)	(87.74)	(80.29)
/t/	1240.06	1184.48	957.12	778.61
	(199.08)	(129.49)	(125.23)	(161.79)
	Values in Bark			
/d/	7.72	6.83	5.95	5.28
	(1.01)	(0.50)	(0.65)	(0.66)
/t/	9.64	9.47	8.14	6.93
	(1.03)	(0.71)	(0.81)	(1.20)
	Values in Log			
/d/	6.77	6.62	6.46	6.32
	(0.16)	(0.09)	(0.12)	(0.14)
/t/	7.06	7.04	6.84	6.63
	(0.17)	(0.13)	(0.13)	(0.20)

Table 10

Group means and standard deviations for fundamental frequency (F0) at onset of the postconsonantal vowel across speaker age and between voicing conditions.

	Speaker Age-Level			
	2.5 Years	4.5 Years	10 Years	Adult
		Values in Hertz		
/d/	319.40	294.60	235.21	147.59
	(40.93)	(31.41)	(29.96)	(44.96)
/t/	313.72	299.92	247.78	163.42
	(43.19)	(28.57)	(25.36)	(52.12)
		Values in Bark		
/d/	3.09	2.87	2.31	1.45
	(0.38)	(0.30)	(0.29)	(0.44)
/t/	3.04	2.92	2.42	1.61
	(0.40)	(0.27)	(0.24)	(0.51)
		Values in Log		
/d/	5.75	5.68	5.45	4.94
	(0.12)	(0.10)	(0.13)	(0.32)
/t/	5.74	5.69	5.50	5.04
	(0.13)	(0.10)	(0.14)	(0.33)

Table 11

Group means and standard deviations for ratios of plosive burst amplitude relative to vowel amplitude (in volts) across speaker age and between voicing condition.

	Speaker Age-Level			
	2.5 Years	4.5 Years	10 Years	Adult
/d/	1.06	0.82	1.04	0.90
	(0.37)	(0.24)	(0.35)	(0.33)
/t/	1.66	1.45	1.33	1.17
	(0.65)	(0.54)	(0.46)	(0.54)

postconsonantal vowel were higher for /t/ than for /d/ across all speaker age groups.

Variability for F1 measures was high across all speaker age groups; standard deviations (in Hz) across speaker age groups ranged from 67.48 to 199.08. Generally, variability for F1 was less on production of /d/ than of /t/. The 2.5 year old speakers displayed the greatest variability for F1 values in both voicing conditions. The least variability for F1 values associated with production of /d/ occurred for 4.5 year old speakers; variability on production of /d/ was comparable for 10 year old and adult speakers. On production of /t/, the least variability occurred for 10 year old speakers, followed by 4.5 year old and adult speakers.

Table 10 summarizes means and standard deviations for measures of F0. F0 was marginally higher for /t/ than for /d/ across all speaker age groups with the exception of the youngest. Differences in F0 between /t/ and /d/ were greatest for adult speakers. Variability was greatest for adults, probably reflecting the broad range of F0 values that resulted from combining data from male and female speakers. Across the children's data, variability increased as speaker age decreased.

The data regarding plosive burst amplitude are summarized in Table 11. Plosive burst amplitude ratios were higher for /t/ than for /d/ in all speaker age groups.

Variability was comparable among all age groups and greater variability was noted for production of /t/ than for /d/.

In order that statistical comparisons could be made among the various speaker age groups, values obtained from only one of the two spectral normalization transformations were included in subsequent statistical analyses. Three 2-way univariate analyses of variance with repeated measures on the factor of voicing were performed to decide which values to use. Independent variables included speaker age group and voicing condition; separate analyses were conducted using either the frequency data in Hertz, in Bark or in natural logarithm. The results of the analyses were similar regardless of which value of spectral measures was used. Significant differences were found for measures of F1 among speaker age groups and between voicing conditions. For measures of F0, significant differences were found only among speaker age groups. Specific effects are described below.

First formant: The 2.5 year old and 4.5 year old speakers differed from 10 year old and adult speakers in F1 measured in Hz (*F* [3, 76] = 20.625, *p* < .0001), F1 measured in Bark (*F* [3, 76] = 26.844, *p* < .0001), and F1 measured in log (*F* [3, 76] = 29.803, *p* < .0001). F1 values were higher for /t/ than for /d/ measured in Hz (*F* [1, 76] = 58.151, *p* .0001), measured in Bark (*F* [1, 76] = 96.976, *p* < .0001), and measured in log (*F* [1, 76] = 111.691, *p* < .0001).

Fundamental frequency: Regarding measures of F0, 2.5 year old and 4.5 year old speakers differed from 10 year old and adult speakers in F0 measured in Hz (F [3, 76] = 63.137, p < .0001), measured in Bark (F [3, 76] = 63.236, p <.0001), and measured in log (F [3, 76] = 54.254, p < .0001). A main effect for voicing condition on F0 was not statistically significant, nor was an interaction of speaker age and voicing condition.

As the results from the statistical analyses were comparable, only values transformed to the log scale were included in subsequent statistical analyses. This scale was chosen as it transformed values along a logarithmic scale throughout the frequency range of interest. The formula used to transform spectral values to the Bark scale in this study did not include a low-frequency correction factor. It was initially proposed that the Bark scale would be used to transform the spectral measures for inclusion in the statistical analyses. Use of the low-frequency correction factor described by Syrdal and Gopal (1986) was thought to be appropriate given that F0 and F1 were the measures of interest. However, the correction factor did not treat F0 values of interest in a sensitive enough manner. Thus, a natural log transformation of the F0 and F1 measures was used to compare differences across speaker age groups (Nearey, 1992).

A 2 factor (4 x 2) multivariate design with repeated measures on the factor of voicing condition explored whether the differences observed in the acoustic data relative to voicing condition and speaker age were statistically significant. The independent variables included (1) speaker age group with fcur levels and (2) voicing condition with two levels. The four dependent variables included VOT, F1 (in log) at onset of the postconsonantal vowel, F0 (in log) at onset of the postconsonantal vowel, and plosive burst amplitude relative to the amplitude of the postconsonantal vowel. Results of the MANOVA indicated that statistically significant differences in the acoustic variables existed among the speaker age-groups (F [3, 76] = 4.04, p = .01). Similarly, statistically significant differences of the acoustic variables were found to exist between voicing conditions (F [7, 532] = 432.45, p = .000). Finally, the interaction of speaker age and voicing condition was found to be statistically significant (F [21, 532] = 3.59, p = .0007).

Multiple univariate analyses of variance were conducted to identify the sources of significant effects identified in the MANOVA. Figures 11 to 14 graph the results from the univariate analyses for measures of VOT, F1 (log), F0 (log), and plosive burst amplitude, respectively. As described previously, only one perceptually validated token of 'tot' and 'dot' spoken by each subject were included in the

Figure 11

Mean VOT values (in ms) of perceptually-validated tokens of /dɑt/ and /tɑt/ across speaker age and between voicing conditions.

		Speaker Age-Level			
Voicing		2.5 Years	4.5 Years	10 Years	Adult
/d/	M.	12.57	14.90	20.57	-11.69
	S.D.	(12.14)	(13.44)	(5.50)	(48.96)
/t/	M.	87.13	98.55	84.04	88.65
	S.D.	(38.35)	(24.91)	(16.77)	(14.20)

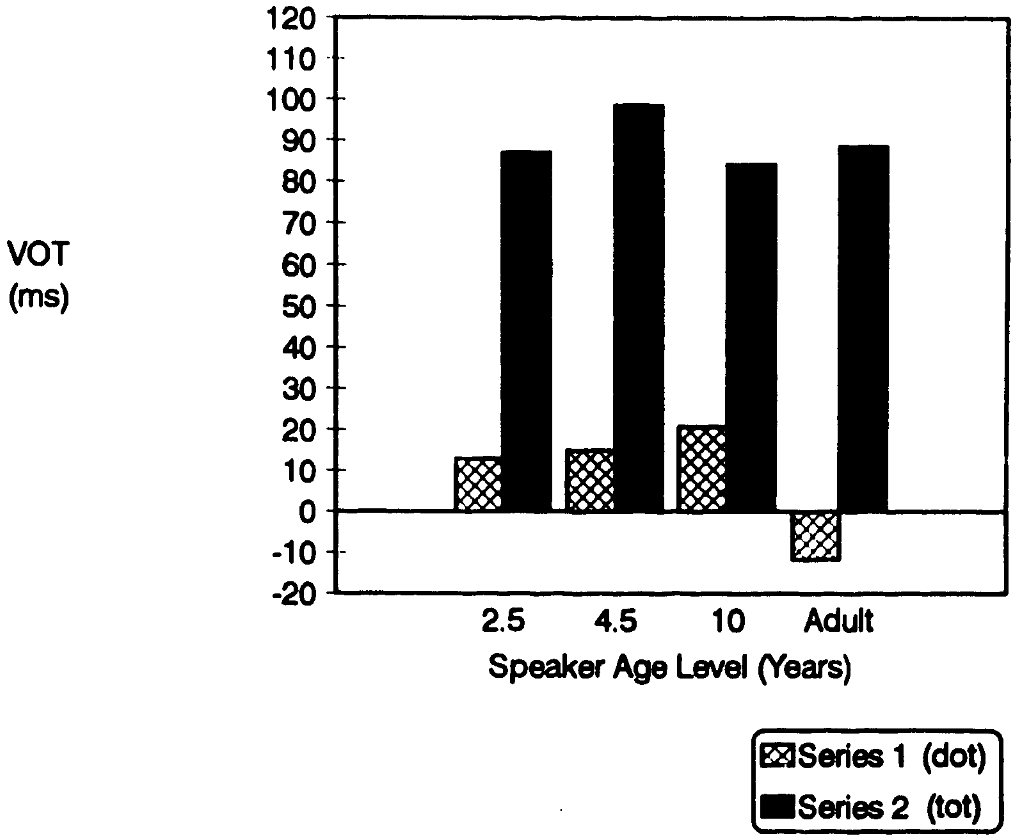


Figure 12

Mean F1 values (in log) of perceptually-validated tokens of /dɑt/ and /tɑt/ across speaker age and between voicing conditions.

		Speaker Age-Level			
Voicing		**2.5 Years**	**4.5 Years**	**10 Years**	**Adult**
/d/	M.	6.80	6.59	6.43	6.30
	S.D.	(0.30)	(0.10)	(0.12)	(0.15)
/t/	M.	7.00	7.04	6.81	6.63
	S.D.	(0.32)	(0.17)	(0.21)	(0.18)

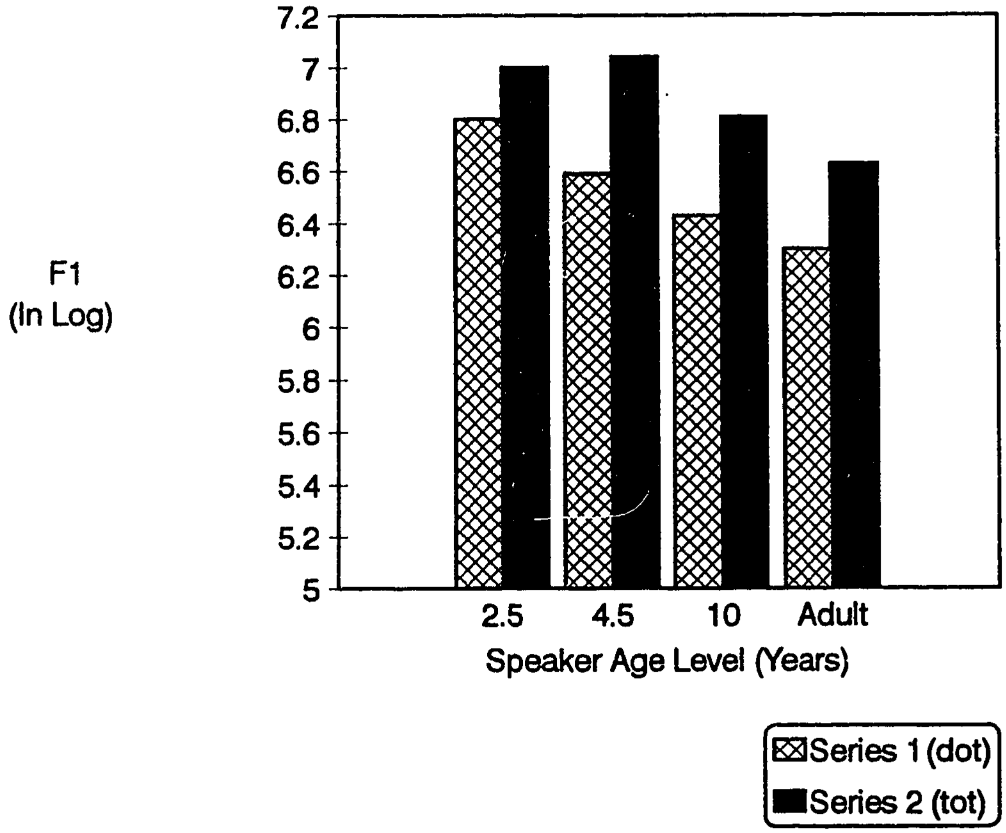


Figure 13

Mean F0 values (in log) of perceptually-validated tokens of /dɑt/ and /tɑt/ across speaker age and between voicing conditions.

		Speaker Age-Level			
Voicing		**2.5 Years**	**4.5 Years**	**10 Years**	**Adult**
/d/	**M.**	5.74	5.66	5.46	4.98
	S.D.	(0.18)	(0.11)	(0.14)	(0.36)
/t/	**M.**	5.70	5.70	5.49	5.04
	S.D.	(0.13)	(0.14)	(0.10)	(0.34)

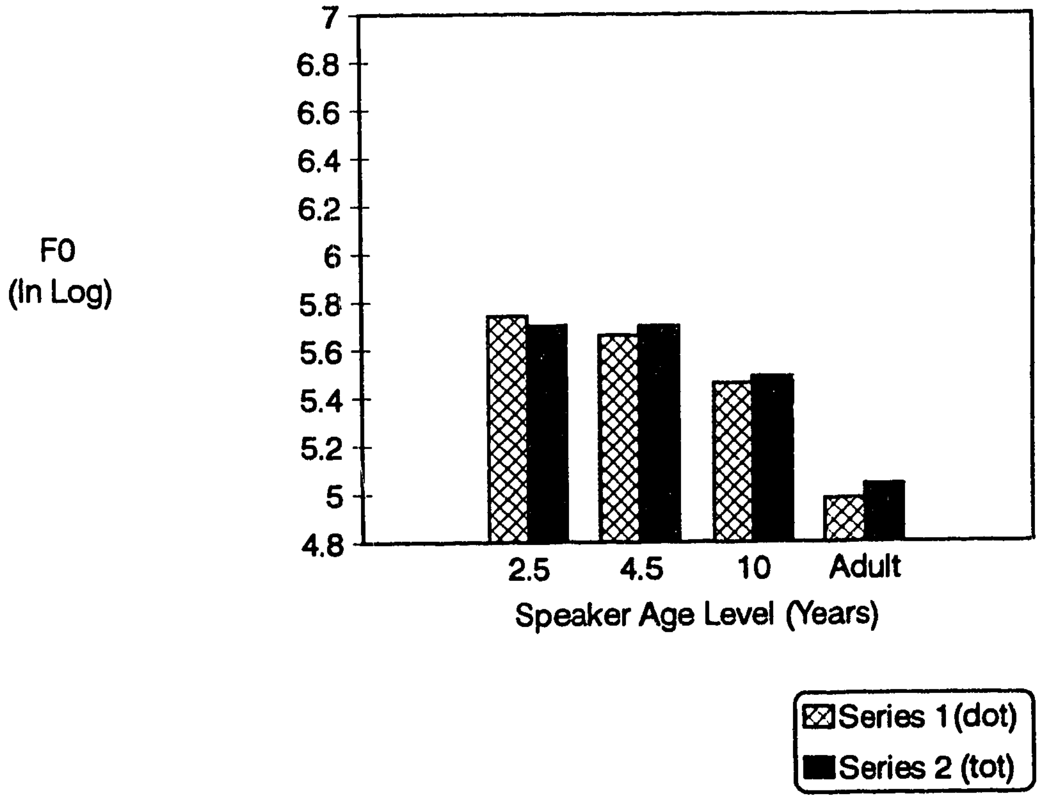
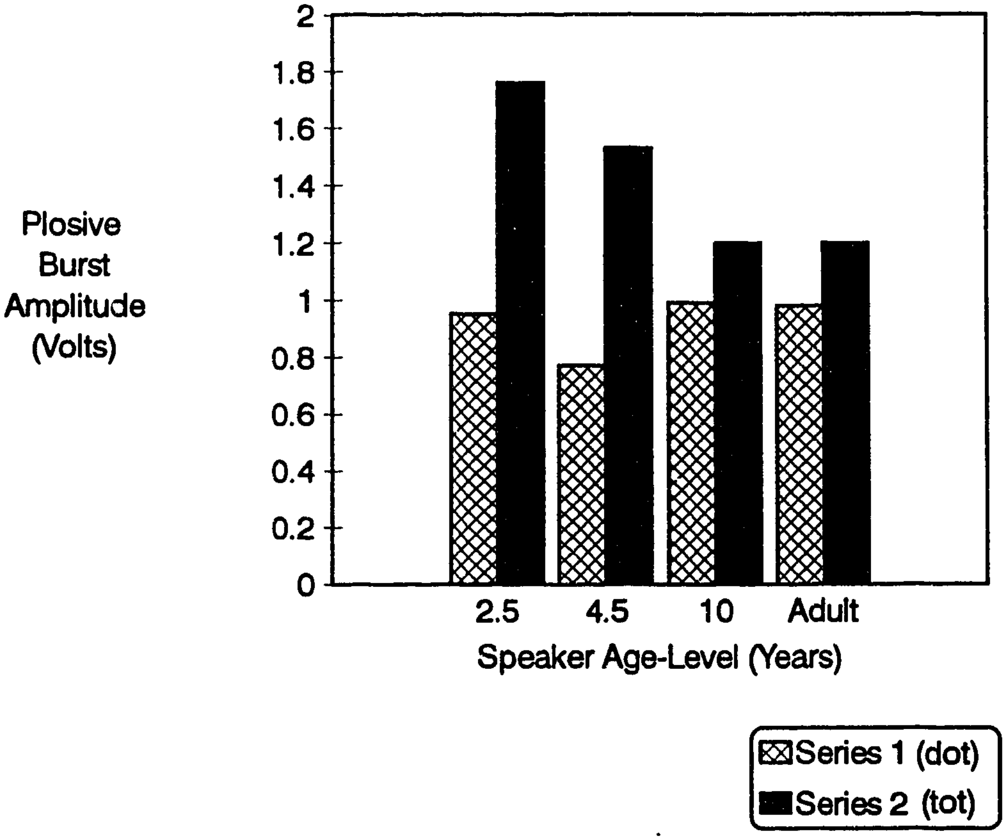


Figure 14

Mean plosive burst amplitude (expressed as a ratio of plosive burst amplitude to vowel amplitude, in volts) of perceptually-validated tokens of /dɑt/ and /tɑt/ across speaker age and between voicing conditions.

Speaker Age-Level

Voicing		2.5 Years	4.5 Years	10 Years	Adult
/d/	M.	0.95	0.77	0.99	0.98
	S.D.	(0.54)	(0.38)	(0.36)	(0.58)
/t/	M.	1.76	1.53	1.20	1.20
	S.D.	(0.59)	(0.62)	(0.44)	(0.81)



acoustical analyses. The mean and standard deviations for these tokens according to each age group and voicing condition are provided in tables preceding each graph.

VOT (Figure 11): The results of the post-hoc ANOVA for measures of VOT revealed that VOT was significantly different between voicing conditions, with longer values associated with production of /t/ (*F* [1, 76] = 387.92, *p* < .00001). A main effect of speaker age (*p* =.02) and an interaction of speaker age and voicing condition (p = .02) approached but did not achieve significance.

First formant (Figure 12): Significant differences in F1 (measured in log) were found to exist among speaker age groups (*F* [3, 76] = 32.13, *p* < .00001) and between voicing conditions (*F* [1, 76] = 120.96, *p* < .00001). Specifically, F1 values for postconsonantal vowels produced by the 10 year old and adult speakers were lower than those produced by the 2.5 year old and 4.5 year old speakers. Also, F1 values were lower for productions of /d/ than of /t/ for all speakers. The interaction of speaker age and voicing condition approached but did not achieve significance (*p* = .04).

Fundamental frequency (Figure 13): Significant differences in F0 (measured in log) were found to exist only among speaker age-groups (*F* [3, 76] = 53.59, *p* < .00001). Specifically, F0 for postconsonantal vowels produced by the adults were lower than those produced by any of the children. F0 for the 10 year old children was lower than that for the 2.5 year old and 4.5 year old children.

Burst Amplitude (Figure 14): Finally, significant differences in plosive burst amplitude ratios were indicated by a main effect for voicing condition ($\underline{F}$ [1, 76] = 49.66, $\underline{p}$ < .00001). Specifically, burst amplitude was significantly higher for /t/ than for /d/. The interaction of speaker age and voicing condition ($\underline{F}$ [3, 76], $\underline{p}$ = .002) also was statistically significant. Burst amplitude was significantly higher for /t/ than for /d/ for 2.5 year old and 4.5 year old children. Additionally, burst amplitude for /t/ produced by 2.5 year old children was higher than for /t/ produced by any other age group.

Discriminant Function Analysis

The remaining experimental questions, which explored the relative contribution of the acoustic cues associated with the voicing contrast to listeners' perceptions of word-initial alveolar plosives, were addressed through a descriptive correlational design using a simple discriminant function analysis. One discriminant function analysis was conducted with one perceptually-validated token from each of 80 subjects included in the original analysis. Equal numbers of 'dot' and 'tot' were represented (i.e., 40 productions of 'dot' and 40 productions of 'tot'). As mentioned previously, a second cross-validation sample of 80 words, one word from each speaker-subject, was used to test the classification accuracy of the discriminant function prediction equation.

Group means for the three predictor variables followed expected trends; that is, production of /d/ was associated with lower VOT values, smaller spectral differences, and lower plosive burst amplitude ratios than production of /t/. Variability for all variables was higher for /t/ than for /d/.

The analysis chosen involved stepwise entry of variables based on minimizing Wilk's lambda. Initially, VOT was entered ($\underline{\lambda}$= 0.286, $\underline{F}$ to enter = 194.78), followed by spectral difference ($\underline{\lambda}$= 0.244, $\underline{F}$ to enter = 13.06), and finally plosive burst amplitude ratio ($\underline{\lambda}$= 0.239, F to enter = 1.76). Wilk's lambda for VOT, spectral difference, and burst amplitude ratio was significant at .0000.

The discriminant function prediction equation was as follows:

$$-4.04 + 0.03\ (\text{VOT}) + 1.61\ (\text{spectral}) + 0.29\ (\text{amplitude})$$

The canonical correlation coefficient of 0.872 indicated a high degree of association between the predictor variables and the criterion groups. The Wilk's lambda of 0.239 suggested that only approximately 24% of variation in the discriminant space was not accounted for by between-group differences.

Wilk's lamda was converted to the chi-square distribution to test the significance of the discriminant function. The ability of the discriminant function to discriminate between /d/ and /t/ was statistically significant, χ^2 (3) = 109.51, p = .0000. The standardized canonical discriminant function coefficients were 0.906, 0.434, and 0.173 for VOT , spectral difference, and burst amplitude ratio, respectively. This suggested that VOT provided the largest contribution to the prediction of voicing between word-initial /d/ and/t/.

Prior to the classification phase of the analysis, the Box's M Test for homogeneity of variance was applied. The covariance matrices between the two samples were significantly different (F [6, 44080.3] = 2.2287, p = .04). The investigator decided to proceed with the classification phase using linear classification rules as opposed to quadratic classification rules for several reasons. Firstly, the SPSS-X computer program for statistical analysis did not allow quadratic classification. Additionally, quadratic procedures perform classification functions poorly for studies of small sample sizes (Marks & Dunn, 1974 cited in Lachenbruch, 1975). Lachenbruch (1975) demonstrated that the discriminant analysis is fairly robust and can tolerate some deviations of assumptions, particularly for studies with equal sample sizes. Finally, the high percentage of correct classifications (discussed later in this section) suggests that the violation of the assumption was not overly detrimental to the analysis (Klecka, 1980).

The classification procedure for the original tokens used to derive the discriminant function prediction equation indicated that 96.25% were correctly classified; only 3 of 40 tokens of /t/ were incorrectly classified. The classification procedure for the cross-validation sample indicated that 98.75% of the sample was correctly classified. For the cross-validation sample, only 1 of 40 tokens of /d/ was incorrectly classified. The high degree of accuracy in classification for both samples suggests a high degree of consistency in the classification function for perceptually-validated productions of word-initial /d/ and /t/.

Discussion

There are several acoustic cues that influence how listeners perceive word-initial plosives with respect to the feature of voicing. The relationship between these cues and listeners' perceptions of the voicing contrast is complex. Much of the information regarding the relationship of these cues with consonant voicing of English plosives has emerged from studies of adult speech. However, there continues to be debate regarding the relative contribution of these cues, individually or in combination, to voicing perception. Even less information is available regarding children's production of these cues and the way in which these cues influence the listener's perception of the voicing contrast.

The purpose of this study was to examine the individual and combined effects of selected acoustic cues on listeners' perceptions of the voicing contrast in word-initial alveolar plosives produced by normal English speakers. Specifically, this study attempted to document differences in knowledgeable listeners' abilities to accurately perceive voicing contrasts in word-initial alveolar plosives produced by speakers of various ages. Secondly, this study attempted to document differences that existed in the various acoustic cues between voicing conditions and among speaker age groups. Finally, it attempted to determine the relative importance of the acoustic cues to the perception of the voicing contrast in word-initial alveolar plosives.

Data were obtained on four acoustical characteristics of word-initial alveolar plosives in single words produced by English speakers at different ages from childhood to adulthood. Acoustic cues thought to differentiate voiced from voiceless word-initial English plosives were targeted for investigation: VOT, F1 at vowel onset, F0 at vowel onset, and plosive burst amplitude. Speakers at four age levels participated in the study: 2 years 6 months to 3 years; 4 years 6 months to 5 years; 10 years to 11 years; and adults. The investigator presumed that these age levels could reflect the progression of acquisition of the voicing contrast. The speakers' taped utterances were digitized on a microcomputer to enable measurement of the acoustic variables. A panel of 5 trained adult listeners judged the word-initial plosives on the digital version of the speakers' recordings as voiced, voiceless, or ambiguous with respect to voicing.

The discussion is divided into four sections. The first section summarizes the results from this study and interprets them in relation to previous studies of the voicing contrast. The second section summarizes whether experimental hypotheses were supported or rejected. The third section summarizes possible threats to internal and external validity, and the final section proposes directions for future study.

Findings and Implications

Acoustic Analysis

Spectral normalization

In order to compare spectral values among speakers with varying vocal tract configurations and sizes, a spectral normalization procedure was required to transform raw data (in Hz) to a normalized scale. Originally, the investigator had proposed to use the Bark transformation. The advantage of using the Bark transformation is that it allows for comparisons of differences in individual formants. A disadvantage relates to the manner in which the Bark transformation treats the data; of concern was the linear representation in the Bark scale of low-frequency data. As discussed previously, differences between logarithmic and Bark transformations are minimal in the mid- and high-frequency ranges. Differences in the low-frequency ranges may be larger using the Bark transformation than could be expected to occur using a logarithmic transformation (Miller, 1989; Nearey, 1989).

A low-frequency correction factor had been described by Syrdal and Gopal (1986). Frequencies below 150 Hz were raised to a value of 150 Hz; separate correction factors were included in the adjusted Bark formula for frequencies between 150 Hz (and lower) and 200 Hz, and frequencies between 200 Hz and 250 Hz. This did not necessarily accommodate this research because the correction factors were not sensitive to frequency values below 150 Hz which could have impacted on the statistical treatment of low-frequency F0 data from adult male speakers. Additionally, Nearey (1989) argued that the proposed correction factor was not supported by psychophysical theory, and was in direct conflict with revisions to the ERB-rate scale proposed by Moore and Glasberg (1983).

In order to test for differences between the normalization procedures, analyses of variance were conducted comparing differences in F1 or F0 between voicing conditions and among speaker age groups using data expressed in Hz, Bark or log. Interestingly, the results were similar regardless of the scale used for both F1 and F0. Age differences existed for both F1 and F0, representing differences in vocal tract size and structure. F1 was significantly higher for /t/ than for /d/ in any scale, but this voicing effect was not found for F0.

The investigator chose to include log transformed values of F0 and F1 in subsequent statistical analyses because the log transformation treated all frequency values in a similar manner and did not require the inclusion of a correction factor. Log normalized values of F1 or F0 were included in the analyses of variance whereas log normalized difference values between F1 and F0 (measured in the same utterance) were included in the discriminant function analysis.

Voice onset time

As predicted by previous findings from studies of the voicing contrast, significant differences in VOT were found between voicing conditions in this investigation, with longer VOT values obtained for the production of voiceless plosives across all age levels. However, significant differences for VOT were not found among speaker age groups or for an interaction of speaker age and voicing condition. Obviously, even the youngest speakers had acquired the voicing contrast, although the large variability among the VOT values for the 2.5 year old children suggested that their production of the voicing contrast was not yet stable. This study supported the finding by Zlatin and Koenigsknecht (1976) that many English-speaking children acquire the voicing contrast by the age of 2.5 years.

In comparison to the VOT data reported by Eguchi and Hirsh (1969) for /t/ in the production of /tɑl/, VOT values in the present study were 8 - 10 ms longer for all ages, but variability was similar. This duration difference may be explained by the different contexts in which the /t/ was produced. Eguchi and Hirsh elicited the words in the sentence "I am tall"; VOT values in a sentence context may be shorter than those obtained on production of isolated words (Lisker & Abramson, 1967; Klatt, 1975). VOT values from the present study were also compared to those reported by Zlatin and Koenigsknecht (1976). Their study used production of word-initial plosives produced in single words as in this research; however, the postconsonantal vowel in their utterances differed from the one employed here (i.e., "dime" and "time" vs. "dot" and "tot"). VOT values for 2 year old children in Zlatin and Koenigsknecht's work were marginally shorter (11.47 ms for /d/ and 69.17 ms for /t/) than those obtained for 2.5 year old children in the present study (17.72 ms for /d/ and 82.83 ms for /t/); however, variability was comparable. VOT values for adult speakers were similar in both studies (i.e., Zlatin & Koenigsknecht - -5.20 ms for /d/ and 87.12 ms for /t/; present study - - 2.23 ms for /d/ and 87.63 ms for /t/); however, less variability in VOT was observed for the adult group in the present study.

Generally, VOT variability decreased with increasing age, suggesting that speakers' productions of word-initial /t/ and /d/ became more consistent as they matured. There was one exception to this trend: Adults showed the greatest variability for VOT values associated with production of /d/ among the speaker groups. This trend towards large variability in VOT on adults' productions of /d/ is also apparent in the data of Zlatin and Koenigsknecht (1976). One possible explanation is the frequent occurrence of prevoicing on adults' productions of /d/ which affects the VOT range. In the present investigation, 29% of productions of /d/ produced by adults were prevoiced compared to 7% of

2.5 year old children's productions, 3% of 4.5 year old children's productions, and 1% of 10 year old children's productions. However, most of the instances of prevoicing were produced by a small number of speakers. Instances of prevoicing were recorded for four 2.5 year olds, two 4.5 year olds, one 10 year old, and eight adults. Consistent with data previously reported in a study by Kewley-Port and Preston (1974), children in this study demonstrated few instances of prevoicing on productions of voiced plosives .

Kewley-Port and Preston suggested that voicing lead may be more difficult to produce than voicing lag. However, it is of interest to note that among the children in the present study, the 2.5 year olds produced the greatest number of prevoiced tokens compared to either the 4.5 or 10 year old children. Additionally, Bond and Wilson (1980) reported that the language-delayed children in their study produced more instances of prevoicing than did the normally-developing children. Zlatin and Koenigsknecht (1976) hypothesized that the infrequent and unstable occurrences of prevoicing in combination with exaggerated voicing lead observed for the youngest speakers reflected their exploration of the phonetic space and instability in the control of temporal gestures of the larynx and supraglottal articulators (pg. 107). It may be that the presence of voicing lead by the adult speakers resulted from well-developed control of the voicing contrast, as suggested by

Kewley-Port and Preston (1974), whereas the wide range of VOT values, including voicing lead, observed in the data of the 2.5 year old children resulted from a lack of such control and consistency.

First formant

Statistical analysis indicated that F1 values (in log) at onset of the postconsonantal vowel were significantly higher on production of /t/ than of /d/. Additionally, all F1 values were significantly higher for the 2.5 and 4.5 year old speakers than for the 10 year old and adult speakers. An interaction between speaker age and voicing condition was not statistically significant.

F1 values at the onset of the postconsonantal vowel /ɑ/ in the utterance 'tot' obtained in the present study were compared to steady-state F1 values for the vowel /ɑ/ reported in other studies. It is acknowledged, however, that F1 values obtained at the onset of the postconsonantal vowel could be expected to be marginally lower than steady-state F1 values as the F1 transition may not be fully complete at voice onset following the voiceless plosive. Table 12 summarizes the mean F1 values (in Hz) and standard deviations (in parentheses) observed in this study compared to those reported by Hodge (1989) and Peterson and Barney (1952). Comparisons with other studies showed that F1 values for the postconsonantal vowel /ɑ/ following the voiceless alveolar plosive /t/ in the present study,

Table 12

Mean F1 values (in Hz) and standard deviations (in parentheses) for /ɑ/ produced by speakers of comparable ages reported in various studies

	Lucky (1992)	Hodge (1989)	Peterson & Barney (1952)
Speaker Age (yrs)			
2.5	1240 (199)		
3		1072 (110)	
4.5	1184 (129)		
5		1010 (84)	1030
9		919 (58)	
10	957 (125)		
Adult	779 (162)	693 (60)	790

(NOTE: Speaker ages represented in the children's data reported by Peterson and Barney [1952] and standard deviations were not specified.)

particularly for the 2.5 and 4.5 year old speakers, were generally higher than those reported in the literature. F1 values reported in Hodge's (1989) study were generally lower than those in the present study. Mean F1 values for children's productions of /ɑ/ reported by Peterson and Barney (1952) were also lower than those observed in the present study and comparable to those reported by Hodge (1989).

An average F1 value of 790 Hz (obtained by averaging the mean F1 values reported for men and women) in the Peterson and Barney study was comparable to that observed for adults in the present study. F1 values for the oldest two age groups were generally comparable to Hodge's data (1989) although variability was greater in the present study. The differences for 10 year old and adult speakers may be partially explained by the fact that Hodge's adult data included F1 values for male speakers only, while the adult data in the present study represented F1 values for both male and female speakers. This combination would be expected to inflate the F1 range and influence the group mean and standard deviation.

Posthoc calculations of the means and standard deviations for the acoustic variables were made according to gender in addition to speaker age and voicing condition to confirm this hypothesis (Appendices E - H). The means and standard deviations for 10 year old males in the present study were actually larger than for the combined data of 10 year old males and females. However, the means and standard deviations were markedly lower for the data of adult males alone than those for the combined data of adult males and females. The mean of 634 Hz (S.D. 28) obtained for males in the present study was comparable to the mean of 693 Hz (S.D. 60) reported in Hodge's data (1989).

The high F1 values and marked variability, particularly for the youngest speakers, is of concern to the present investigator. Young children's speech is expected to be more variable than that of older speakers; however, the mean F1 values and variability across age groups in this study were very different from those reported in Hodge's study. The high values and wide range of F1 observed in this study overlap with values in the range associated with F2 for production of /ɑ/. Measurement error could account for the disparity. The investigator was inexperienced in acoustic measurement. However, the high intra-rater reliability suggests that the results obtained are reliable if not valid.

The investigator noted wide fluctuations in F1 values (not reported in this study) dependent upon the length of the analyzing window and whether FFT or LPC estimates were used, particularly for the 2.5 and 4.5 year old speakers. As Keller (1992) points out, LPC formant estimation can often deviate greatly from values obtained from FFT analysis. However, wide-band analysis may not distinguish relatively closely-spaced formants, such as F1 and F2 for the vowel /ɑ/.

In the present investigation, F1 frequency values were obtained with reference to FFT and LPC analyses of just less than one-half of the first pitch period of the postconsonantal vowel (i.e., effective analysis bandwidth = 2 x the speaker's fundamental frequency) (Kent & Read, 1992). This procedure was guided by suggestions from Milenkovic (1989) and Keller (1992) that isolating some fraction of the pitch period will simulate the effects of wide-band spectral analysis necessary for optimal spectral measurements of speech produced by subjects with higher fundamental frequencies. However, this analysis window may have been more appropriate for LPC analysis. It may be that the wide-band analysis, simulated by the narrow time window used, was too large and did not provide effective resolution of the harmonic frequencies for the younger speakers for FFT - LPC correspondence.

Additionally, variations in intonation, resulting in fluctuating pitch and loudness characteristics, may have influenced the results. This will be discussed in greater detail in a later section.

<u>Fundamental frequency</u>

Results from the ANOVA indicated that significant differences of F0 (in log) existed among the age groups. Specifically, F0 values for adults were significantly lower than those for children. Among the children, mean F0 for 10 year old children was significantly lower than that for the 2.5 and 4.5 year old children. These results were expected given what is known about the developmental changes in F0 associated with increasing vocal tract length and growth of the laryngeal structures (Kent, 1976).

Significant differences for F0 between voicing conditions were not found. Results of the present study were similar to those described by Revoile, Pickett, Holden-Pitt, Talkin, and Brandt (1987) in which only marginal differences in F0 were found between vowels at voice onset following voiced and voiceless word-initial plosives. Results of the current investigation did not support the findings of significant differences in F0 between voiced and voiceless word-initial plosives reported in other studies for adults (House & Fairbanks, 1953; Lehiste & Peterson, 1961; Ohde, 1984), and for 8 and 9 year old children (Ohde, 1985).

A trend existed for higher F0 values to be associated with /t/ than /d/ for the three oldest speaker groups, but this difference was not statistically significant. It may be that F0 measures for the first pitch period were susceptible to substantial variability due to instability in the early portion of the vowel. Ohde (1985) observed the greatest variability in F0 values for the first pitch period in the speech of both adults and 9 year old children. Future investigations of F0 differences as a function of voicing should involve measurement of F0 at perhaps the third pitch period, which still occurs near voice onset but may provide a more stable measure of F0 than would be obtained in the first pitch period.

F0 values in the present study were higher for all age groups than those reported by Hodge (1989); however, variability for all age groups except the adults was comparable to that found in her data. Variability of F0 for the adults was highest for production of both /tɑt/ and /dɑt/. Among the children, variability in F0 at the onset of the vowel decreased with increasing age, suggesting improved consistency with maturity.

As discussed previously for the high variability of F1 observed in the adult and 10 year old speakers' data, a probable explanation for the high F0 variance in the adult group is that the data represented both male and female speakers whose F0 values would be noticeably different. The inclusion of female voices of higher F0 in the adult data would result in a higher mean value and larger variance. Posthoc calculations of the means and standard deviations of F0 according to speaker age, voicing condition, and gender supported this hypothesis (Appendix G). Adult females' voices were characterized by markedly higher F0 compared to adult male voices. When the data were analyzed by gender, variability in F0 for both female and male speakers was smaller than that observed in the combined male and female data.

Plosive burst amplitude

As hypothesized, plosive burst amplitude was significantly larger during production of /t/ than of /d/. Additionally, the interaction of speaker age and voicing condition on plosive burst amplitude was statistically significant. Specifically, burst amplitude was significantly larger for production of /t/ than of /d/ for the two youngest speaker groups. In fact, the mean plosive burst amplitude in the 2.5 year old children's productions of /t/ was significantly larger than the mean plosive burst amplitude for all other speaker groups' productions of /t/.

These results contradict those reported by Revoile et al. (1987) in which only marginal differences in burst amplitude were reported between voicing conditions on production of word-initial plosives. Those authors found the presence or absence of aspiration noise to be a more significant cue to the voicing contrast than burst amplitude for both normal-hearing and hearing impaired listeners. However, while the presence of aspiration was an important cue it was not necessary to the perception of the voicing contrast if more salient cues (i.e., VOT or F1 transition) were available.

Aspects of aspiration were not studied in the present investigation as the investigator felt that effects of aspiration would be related to the temporal measure of VOT, rather than a unique cue to voicing. In fact, Revoile et al.(1987) reported that the effects of presence or absence of aspiration on listeners' perceptions of the voicing contrast could be simulated by the introduction of a silent interval between the burst and the onset of voicing for the postconsonantal vowel. This suggested that the duration between the plosive burst and the onset of voicing (i.e., VOT) was of greater importance to the perception of the voicing contrast than was the presence of aspiration.

Klatt (1975) found differences in burst amplitude duration between voiced and voiceless plosives. Based on these findings, Klatt proposed that differences may also exist in plosive burst amplitude, as perceptual loudness is related to both intensity and duration. This hypothesis is supported by the results from the present investigation.

Perceptual Analysis

Intra-listener reliability

Measures of intra-listener reliability were calculated for the number of complete agreements and for the number of within-category agreements for each listener, and the combined reliability scores from the five listeners were statistically analyzed to determine if significant differences existed in the scores dependent upon speaker age and voicing condition. It was hypothesized that reliability scores would be significantly lower for judgments of words produced by 2.5 year old speakers, and that listeners' judgments would become more reliable as speaker age increased. Generally, reliability scores did improve as speaker age increased but these differences were not statistically significant. Listeners' reliability scores were higher for within-category agreements than for complete agreements, particularly for the utterances of the youngest speaker group. However, when one examines the data set for complete agreements, the main effect for differences in reliability scores among speaker age groups approached but did not achieve statistical significance. Statistical significance may have been obtained had more tokens been included in the analysis.

Inter-listener comparisons

The results of the analyses of variance analyzing differences in the numbers of correct, incorrect, and ambiguous judgments generally supported the predictions of the investigator. In summary, listeners made fewer correct judgments, more incorrect judgments, and more ambiguous judgments of plosives produced by 2.5 year old speakers compared to those produced by any other age group. Effect sizes for the perceptual data, based on calculations of effect size for paired comparisons (Kraemer & Thiemann, 1987), were as follows: medium to large effect sizes were obtained for significant differences in the numbers of correct judgments among speaker groups; small to medium effect sizes were obtained for significant differences in the numbers of ambiguous judgments among speaker groups; small effect sizes were obtained for significant differences in the number of incorrect judgments among speaker groups (Appendix I -- Calculation of Effect Sizes).

Generally, with the exception of listener L4 (children), there were relatively high numbers of correct responses for plosives produced by all age groups; however, listeners' accuracy markedly improved as speaker age increased. If the data from L4 are excluded, listeners correctly judged between 69% and 85% of plosives produced by the 2.5 year old speakers. When judging the speech of the 4.5 year old speakers, listeners correctly judged between 88% and 98% of the plosives. For the 10 year old and adult speakers, listeners' accuracy again increased to between 91% and 100%.

Few incorrect responses occurred for any age group. As hypothesized, the largest percentage of incorrect judgments (between 1% and 10% of the plosives produced) was made for plosives produced by the youngest speakers. Listeners rarely made incorrect judgments for plosives produced by the

4.5 year old speakers, and never for the 10 year old and adult speakers.

Ambiguous judgments were made more frequently than incorrect judgments by all speakers. L4 (children) made substantially more ambiguous judgments than did any other listener, particularly for judgments of /d/. However, although the response patterns were different (Figures 7 & 10) dependent upon whether the data from L4 were included, the results of the statistical analyses were similar. As hypothesized, all listeners made more ambiguous judgments for plosives produced by the 2.5 year old children than for those produced by any other age group.

When ambiguous or unreliable judgments occurred for the youngest and oldest speakers, the ambiguity occurred most often on production of voiceless alveolars; that is, some of their productions of voiceless alveolar plosives resembled voiced alveolar plosives. The 4.5 year old and 10 year old children exhibited the opposite pattern; specifically, they produced voiced alveolar plosives that resembled voiceless plosives. This latter trend may be easier explained than the former trend. These children may have been attempting to produce "clear speech" (Kent & Read, 1992). Certainly, the experimental task may have fostered effortful, well-articulated speech production resulting in VOTs and F1 transitions of longer duration, and plosive bursts of larger amplitude (this is discussed in greater detail in a later section).

It seems unusual that the perceptual results for the youngest and oldest speakers were similar. The mechanisms underlying the trends may be quite different. Adults may be more likely to make their speech productions economical, perhaps by reducing the amount of time to produce speech sounds. The 2.5 year old children, though, may only recently have acquired the voicing contrast and may not yet consistently produce obvious voicing distinctions for the listeners.

Obviousl all speakers in this sample could produce a voicing contrast for word-initial alveolar plosives. In fact, by 4.5 [illegible], speakers demonstrated consistently clear and unambiguous productions of the voicing contrast in word-initial alveolar plosives. Even the alveolar plosives produced by the 2.5 year old speakers were frequently perceived correctly. However, listeners' responses implied that the youngest speakers were inconsistent in their productions. This assumption was supported by evaluation of the acoustic data. Although the mean values generally fell within expected limits, the standard deviations indicated substantial variability. It would appear that the variability, and thus overlap of voicing boundaries, in the acoustic cues resulted in increased uncertainty for the listeners.

Analysis of the acoustic variables also revealed a high degree of variability for adult speakers, particularly for the spectral variables. This was accounted for by the inclusion of values from both male and female speakers, however. Given the high degree of accuracy in listeners' perceptions of plosives produced by the adult speakers, it appeared that the gender differences that influenced the central tendency of the adults' frequency data did not influence listeners' perceptions of the voicing contrast in those data.

As mentioned previously, perceptual results from the "fourth" listener included in the statistical analyses were actually a combination of the results from two listeners (one who judged all of the children's data and one who judged the adults' data). L4 (children) made noticeably more ambiguous judgments, particularly for judgments of /d/, than did the other listeners. Similarly, her reliability scores for complete agreements were lower than these scores for other listeners, particularly for judgments of /d/. It appeared that she used the ambiguous categories ('t?' and 'd?') more frequently than did the other listeners. Anecdotal information suggested that L4 may have interpreted the perception task differently from the other listeners. Instead of making only a perceptual judgment of the word-initial plosive (i.e., reporting what she heard), she may have been making qualitative judgments as to the "goodness" of the plosive. Although identical instructions were given to all listeners, the experimenter's description of the task may not have been sufficient to permit all listeners to respond in a similar fashion. This will be discussed in greater detail in a later section.

Most studies of voicing contrasts in children's speech have not included strong perceptual-validation components. However, the present results support the trends observed by Hodge (1989) and Peterson and Barney (1952) for perceptual judgments of speakers' productions of vowels. Peterson and Barney (1952) reported that reliability of listeners' judgments of vowels was lower for children's productions than for adults' productions. Hodge (1989) found that listeners' accuracy in identifying various vowels increased with speaker age. In her study, both listener accuracy and reliability improved with speaker age. Her perceptual results were supported by acoustic evidence that younger speakers' productions of acoustic cues were less well-defined and more prone to confusion than were older speakers' productions. Similarly, in the present study, the acoustic results for the youngest speakers revealed substantial variability and overlap which were hypothesized to result in increased incorrect and ambiguous judgments of the voicing contrast by the listeners.

Contribution of the Acoustic Cues to Listeners' Perceptions

Due to limitations in group sample size, the data from this study could not be analyzed for information regarding the influence of speaker age on the contribution of the various acoustic cues to the perception of the voicing contrast of word-initial alveolar plosives. Some results were obtained from the discriminant function analysis, however, regarding the relative importance of each of the acoustic cues to the perception of voicing across speaker age, and the accuracy with which perceptually-validated word-initial alveolar plosives could be correctly classified as either voiced or voiceless based on the acoustic information provided. The discriminant function analysis of the acoustical data was limited to tokens of /t/ and /d/ that were correctly perceived by 4 out of 5 listeners. Based on these perceptually distinct tokens, the capacity of the discriminant function to predict voicing category was high. Ninety-six percent of the original sample (N = 80 tokens) and 99% of the cross-validation sample (N = 80 tokens) were correctly classified. VOT provided the greatest information regarding voicing category, followed by spectral differences (F1 - F0), and lastly plosive burst amplitude. The experimenter's reason for including only perceptually distinct tokens was to facilitate the generation of the best possible prediction equation with the specified acoustic parameters. The strong prediction equation generated by these clear tokens indicated that the acoustic cues provided substantial information to listeners enabling them to correctly distinguish the voicing contrast. This is particularly interesting given that high classification rates were obtained without the aid of other contextual and conversational cues. In conversational speech, the listener has a variety of sources of information to enable accurate perception. Even without this information in the present study, the listeners were able to accurately perceive these tokens.

These results tend to support the argument for VOT as the primary cue to voicing, with F1 onset frequency providing important but secondary information (Lisker, 1975; Summerfield & Haggard, 1977). As discussed previously, Lisker (1975) argued that evidence of context sensitivity rules out F1 as a salient cue for voicing. Simon (1974) (cited in Summerfield & Haggard, 1977) suggested that using F1 as a cue to voicing may be a learned response whereas using the temporal cue of VOT may be inborn. In that study, children younger than 5 years of age correctly perceived voiced velar plosives in the absence of an F1 transition whereas children older than 8 years of age required the presence of a low F1 onset frequency to allow accurate perception. The results from Revoile et al. (1987) also indicated a primary role for VOT in voicing perception of word-initial plosives. F1 transition was found to be an important cue for normal-hearing and hearing-impaired listeners but could be overridden by manipulating VOT to provide conflicting information. In the present study, plosive burst amplitude ratio also distinguished voiced and voiceless plosives but this contribution was not essential to the perception of the voicing contrast.

The classification function generated in this investigation may only be true for clear productions. It may be that listeners use various cues in different ways dependent upon the age of the speaker and the clarity of the production. Perhaps, particularly if contextual cues are not available, listeners rely more heavily on F1 onset frequency, plosive burst amplitude, or other acoustic cues when the acoustic features are less clear. In the data reported by Forrest and Rockman (1988), 39% of correctly perceived voiceless word-initial plosives produced by phonologically disordered speakers had VOTs that overlapped with VOT values of incorrectly perceived plosives. In these cases, it appeared that VOT was not the dominant cue for the perception of the voicing contrast. These authors provided some evidence that other cues contributed to listeners' correct perceptions even when VOT was ambiguous. Similarly, in a study of normally-developing 3 and 4 year old children, Menyuk & Klatt (1974) found that for word-initial plosives produced in singleton or cluster contexts by these children and correctly perceived by knowledgeable listeners (as evidenced by their correct transcriptions of the spoken words), there was no unique VOT boundary between the voiced and voiceless word-initial plosives. Menyuk & Klatt (1974) suggested that the listeners must have used cues other than VOT to enable their accurate perceptions of the voicing contrast. Further study is needed to determine the combination of cues and many conditions that influence the perception of voicing in word-initial plosives.

Experimental Hypotheses Revisited

Generally, the experimental hypotheses conceived at the onset of this work were supported by the results, with some exceptions:

Perceptual results

The experimenter had hypothesized that listeners' correct judgments would increase while listeners' ambiguous and incorrect judgments would decrease as speaker age increased. These hypotheses were supported by the results. Specifically, the results were as follows:

1. The frequency of listeners' correct judgments of word-initial alveolar pl)sives increased as speaker age increased.

2. The frequency of listeners' incorrect judgments of word-initial alveolar plosives decreased as speaker age increased, and the greatest number of incorrect judgments occurred for plosives produced by the youngest speakers.

3. The frequency of listeners' ambiguous judgments of word-initial alveolar plosives decreased as speaker age increased, and the greatest number of ambiguous judgments occurred for plosives produced by the youngest speakers.

Acoustic results

The experimenter had hypothesized that the values of the four acoustic parameters would differ between voiced and voiceless alveolar plosives. Additionally, it was hypothesized that age effects would be found for F1 and F0, with higher values observed for the younger speakers than for the older speakers. Finally, the experimenter predicted that some interaction effects would be statistically significant: It was predicted that VOT values for voiceless alveolar plosives produced by the 2.5 year old children would be shorter than those produced by all other age groups; that the F1 onset frequency differences between voiced and voiceless plosives would be significant only for the two oldest speaker groups; and that F0 onset frequency differences would be significant only for the two oldest speaker groups. The results were as follows:

1. VOT values were significantly longer for production of voiceless alveolar plosives than for voiced alveolar plosives. However, a predicted interaction of speaker age and voicing condition was not statistically significant.

2. F1 frequency at onset of the postconsonantal vowel was significantly higher for production of voiceless alveolar plosives than for voiced alveolar plosives. F1 frequency was higher for the 2.5 and 4.5 year old speakers than for the 10 year old and adult speakers. A predicted interaction of speaker age and voicing condition was not statistically significant, however.

3. F0 frequency at onset of the postconsonantal vowel was significantly higher for the children than for the adults, and for the 2.5 and 4.5 year old children than for the 10 year old children. A prediction of significantly higher F0 values for production of voiceless alveolar plosives than for voiced alveolar plosives was not supported statistically, nor was a predicted interaction of speaker age and voicing condition.

4. Plosive burst amplitude was significantly larger for production of voiceless alveolar plosives than for voiced alveolars for all speaker groups. Plosive burst amplitude was not significantly different among the speaker groups.

These results supported the experimental hypotheses.

The experimenter's predictions for statistically significant interaction effects were based on the assumption that the youngest speakers would not consistently produce distinctive voicing contrasts. It was expected that the 2.5 year old speakers would produce voicing contrasts but that the values for their voiceless plosives would overlap with those of the youngsters' voiced plosives. Additionally, the investigator had predicted that the 2.5 and 4.5 year old speakers' productions of F1 and F0 would be marked by sufficient variability, particularly for the 2.5 year olds, to obfuscate statistically significant differences between voicing conditions for these age groups. These predictions were not supported. In actuality, the 2.5 year old speakers often produced plosives with excessively lengthy VOT values, and large differences in F1 and plosive burst amplitude between productions of voiced and voiceless plosives.

Relationship of acoustic variables to perception

1. The prediction that successful classification of word-initial alveolar plosives would increase as speaker age increased based on the contribution of the acoustic cues could not be tested due to insufficient sample size.

Threats to Validity of the Results

Various conditions and procedures used in this study affected the internal and external validity of the experimental results. Threats to validity have been summarized by many authors. The categories described by Drew and Hardman (1985) have been used as the frame of reference for the following discussion. Generally, confounding effects to internal validity may be categorized in the following manner: threats arising from maturation, test practice, instrumentation, and the Hawthorne effect. Threats to external validity or generalizability are categorized in the following manner: threats arising from sample differences, pretest influence, and treatment restrictions. Each of these will be discussed in turn, and alternative procedures will be considered.

Internal Validity

Effects of maturation

Maturation effects refer to factors such as time or fatigue that may influence a subject's performance on the experimental task. Various maturation effects may have influenced the speakers' and listeners' performances in this investigation. Among the speakers, testing occurred at various times of day according to the preference of the subject or guardian. However, the effect of time of day may have been balanced across speaker groups insofar as subjects within each group were tested at various times. Administration of the screening procedures (i.e., screening procedures for hearing, speech, and language) preceded administration of the experimental task so speakers may have experienced fatigue at this point in the test session. However, as all subjects were exposed to the same testing sequence, differences in test performances among the speaker groups due to fatigue should be minimal. Additionally the experimental task required minimal effort from the speakers as the speaking task involved labelling of a reasonably small number of single-syllable words (n = 32).

Effects of fatigue and boredom on test performance were of greater concern to the investigator for the listeners than for the speakers. The perceptual task was conducted over the course of one day. Each of the five listeners participated in separate test sessions lasting approximately one and one-half hours with one 15-minute break. Testing occurred at various times in the day according to the listener's preference and the examiner's schedule. The listeners heard a total of 920 words, including 10 practice words prior to the presentation of each of four blocks of 220 test words. The task was lengthy and monotonous, thus being susceptible to the effects of listener fatigue and boredom. However, the order of presentation of the four blocks of words (corresponding to the four age groups) was varied among the listeners to prevent marked effects of fatigue from interfering with listeners' perceptions of one particular speaker age group.

One other effect of maturation may relate to the experimental design employed in this study. A cross sectional design was used to assess the developmental progression of the acquisition of the voicing contrast. When a cross-sectional design is employed to chart the progression of a certain parameter across time, the experimental effect cannot be attributed solely to the influence of age; rather, differences in social, educational, and other environmental factors that may have changed across generations could have contributed to the effect. A longitudinal study allows for more confident interpretation of the effect of aging. The investigator chose the cross-sectional design because of expediency. Additionally, the investigator believed that age and neuromuscular maturity, and not other environmental factors, would be the largest factors influencing the development of the voicing contrast if the speakers displayed normal cognitive, speech, and language development.

Effects of test practice

Speakers were required to produce multiple repetitions of the target words. The investigator considered this to be a necessary component given the high intra-speaker variability that exists in production of the acoustic values. Multiple repetitions allowed the investigator to obtain a representative sample of the central tendency and range of values for the acoustic variables that occur within and across speakers. The effects of practice were minimized in two ways. Firstly, as all speakers produced the list of words in the same order, the practice effect should have been balanced. Secondly, the target words were embedded within the word list so that minimal pairs were never presented successively, thus minimizing any effects related to practice or fatigue that may occur from multiple successive repetitions of the same articulatory pattern.

Effects of instrumentation

Effects of instrumentation may have occurred related to changes in the behaviour of the examiner or related to the acoustic analysis system used. The acoustic measurement difficulties have been previously described. Measurement error associated with acoustic measurement of voices of higher fundamental frequencies, particularly children's voices, has been well-documented. As expected, F1 values obtained from acoustic measurement of the target words produced by 2.5 year old children and, to a lesser degree, 4.5 year old children varied greatly. Even when the same word was measured with various analyzing windows, F1 measurements could deviate substantially from one measurement to the next. A dilemma existed for measurement of F1 at the onset of the vowel /ɑ/. A problem with FFT formant estimation from a wide-band analysis is that it may not distinguish the relatively closely spaced F1 and F2 formants. On the other hand, LPC formant estimation may deviate substantially from narrow-band or wide-band measurement, particularly for voices of higher fundamental frequency (Keller, 1992).

Perhaps, the use of a larger time window may have resolved the harmonic frequencies in the children's data more effectively, thus allowing more accurate formant estimation. Alternatively, comparisons could have been made between F1 measures made with CSpeech and other digital speech analysis systems, such as those allowing spectrographic measures, to determine whether large differences in mean values and variability existed between measurement techniques. The results of F1 measurements reported in this investigation need to be interpreted cautiously.

Changes in the examiner's behaviour may also have influenced the acoustic measurement. As mentioned previously, the examiner was inexperienced in acoustic measurements. However, substantial preparation, including the development of a measurement protocol and repeated practice of the acoustic measurements occurred prior to recording the actual acoustic values. It may be that the experimenter's ability to perform the acoustic measures improved over time; however, as the subject's data were measured in random order, this effect should be balanced across speaker groups. Additionally, as the intra-rater reliability was acceptably high, it suggests that the measurement procedures did not deviate substantially.

Perhaps the greatest effect of examiner behaviour occurred regarding the instructions to the listeners. Identical written instructions were displayed and read to the listeners. However, differences in the performances among the listeners suggest that the examiner may not have been explicit enough regarding the task instructions to invoke the same judgment criteria across judges. Specifically, the nature of the task and the use of the ambiguous categories could have been better explained to avoid differences in test performance among the listeners.

The Hawthorne effect

A Hawthorne effect may occur due to changes in subjects' performances related to the artificiality of the experimental situation. Various differences in the subjects' speaking style and in the observed acoustic variables from what may be expected to occur in more natural situations suggest that the experimental task may have influenced the speakers.

The acoustic values observed in this study frequently were of larger degree than those reported in other studies. For example, VOT values were generally longer, plosive burst amplitudes particularly on production of voiceless alveolar plosives were often larger, and F1 and F0 values were often higher than values reported in other studies. These differences may be related to the fact that many speakers in this study produced the target words in an effort to be highly intelligible. Kent and Read (1992) discuss the differences between clear speech, such as that produced by subjects in this experiment, and conversational speech. Clear speech is characterized by a lengthening of consonant and vowel segments, slower speech rate and greater intensity on production of obstruent sounds, whereas conversational speech is marked by modified or reduced forms, quicker speech rate, and economy of production. In clear speech, speakers make an effort to produce speech that is acoustically distinct. Speakers in this task, with the possible exception of the youngest speakers, perceived an importance in producing clearly articulated speech for the experimental task. Thus the acoustic values of their productions may have been exaggerated relative to those of the speakers' natural, conversation speech patterns.

The format of the experimental task may have fostered effortful speech production. Production of isolated single words would be expected to be associated with exaggerated acoustic values in comparison to production of words within sentence or conversational contexts. Additionally, psychological stress inherent in a "test" situation, such as that induced by the introduction of the head-mounted microphone and the tape recording apparatus, could have resulted in effortful speech production. Although the clear speech produced by these speakers does not approximate natural conversation, it may approximate speech patterns obtained in a formal speech-language assessment in which there is a real and perceived need to produce well-articulated, 'best effort' speech.

The single-word production task was chosen to reduce other confounding effects, such as phonetic context, intonational fluctuations, and speaking rate, which are known to influence production of the voicing contrast. The use of the sentence frame in an attempt to control such effects did not eliminate them, however. For example, although the percentage of occurrence of the pattern was not calculated, the investigator noted several instances in which speakers used an interrogative (pitch rising throughout production of the target word) instead of a declarative (pitch falling or remaining even throughout production of the target word) intonation, possibly due to speaker uncertainty about performance accuracy or nervousness about the recording task.

Intonational variations were particularly difficult to control with the youngest speakers. These children were highly variable in their productions of the target words, possibly because of the novelty of the situation in addition to the expected performance variability related to their neuromuscular maturity; therefore, their utterances were often characterized by wide fluctuations in loudness and pitch. The high degree of variability for the acoustic measures obtained for the 2.5 year old speakers may be partly attributable to these differences.

External Validity

Population-sample differences

This dimension of external validity concerns the degree to which the subjects who participated in the study are representative of the population of English speakers. Two subject groups participated in this study: the speakers and the listeners. While socioeconomic status and educational background were not recorded, the speakers were generally from middle-class and professional families. Daycares who agreed to participate in the study were either teaching institutions or affiliated with hospitals or educational institutions. It is likely that children and adults from less educated and lower income families were not equally represented in the sample.

The listeners were all speech-language pathologists; therefore, the perceptual results can only be generalized to listeners knowledgeable about the voicing contrast of interest. Although most normal English-speaking adults would be able to discriminate between /t/ and /d/, experience in evaluating speech may bias the way in which listeners attend and respond to the experimental task.

Pretest influence

Drew and Hardman (1985) argue that a pretest or warm-up procedure may result in two threats to external validity. Firstly, it causes the testing to be artificial and less like a real life setting. Secondly, it may either increase or decrease the subject's sensitivity to the task. Nonetheless, as described previously in the methodology, the investigator chose to utilize a pretest procedure for both the speaker and listener tasks. The primary reason for including the pretest procedures was to familiarize the subjects with the task. The investigator acknowledges that the experimental task was highly artificial.

Treatment restrictions

The greatest restriction to the generalizability of the results relates to the limited articulatory context in which the voicing contrast was studied. The interpretation of the results is limited to isolated word-initial alveolar plosives in one vowel context. Obviously, this does not approximate conversational speech. The investigator sacrificed some level of gencralizability to allow for better internal validity. Specifically, production of isolated words was chosen to facilitate responses by the youngest children. Additionally, this context provided some control over the influence of phonetic environment, increased articulatory load, and differences in stress and intonation. The vowel context was selected to enhance the effect of F1 transition on the voicing contrast. Examination of the acoustic cues indicated that the results were not typical of those for words produced in sentence or conversational contexts. Any conversational or phonetic cues available to the listeners were minimized. Although the interpretations of the results are limited, they suggest some cues that may aid perception of the voicing contrast and point to directions for future investigation.

Directions for Future Study

Although the voicing contrast in speech has been extensively studied, there continue to be questions about the interaction of perception and production in the appreciation and acquisition of the contrast. This study provided information on this interaction in English-speaking children and adults for one context in which the voicing contrast occurs. A number of directions for future research arising from this study are offered.

In order to trace the development of the voicing contrast from emergence to mastery, younger subjects than those included in this study would be required. The 2.5 year old children in the present study had acquired a voicing contrast for the limited context under investigation. Evidence from Macken and Barton's study (1978) suggests that children may acquire the voicing contrast between 18 and 28 months. Therefore, the experimental design could be replicated including an additional speaker group of children aged 18 months. This might yield more discriminating information regarding the developmental changes in the acoustic characteristics associated with the emergence of the voicing contrast.

The experimental design could be replicated varying the place of articulation (i.e., velar and labial plosives) or phonetic contexts (i.e., varying the postconsonantal vowel; investigating plosives in word-final positions or consonant blends), thus providing further information on whether the acoustic and perceptual patterns observed in this study hold in other contexts. Additionally, a comparison of target words produced in sentence contexts with those produced in isolation would provide information on the influence of other contextual factors on the acoustic parameters and how these changes influence the listener.

It would be interesting to repeat the study using untrained listeners in the perception task. The listeners in the present investigation were speech-language pathologists trained to listen for and evaluate slight differences in speech patterns. Therefore, it seems reasonable to assume that they would be more critical in their judgments of the voicing contrast. Alternately, the acoustic cues inherent in the speech signal may be strong enough to enable consistently accurate responses regardless of listeners' experience with the task.

Use of the discriminant function analysis to allow interpretations of the relative contributions of the acoustic and contextual cues to the perception of voicing should be pursued. Specifically, the hypothesis regarding the changing contribution of the acoustic cues dependent upon speaker age to the perception of voicing needs to be empirically tested. Larger sample sizes than those used in the present investigation would be required. Specifically, between 30 and 60 subjects per speaker group are required for separate discriminant function analyses involving three predictor variables (Klecka, 1980). Additionally, the analysis could be replicated with ambiguous and incorrect tokens to determine whether the relative importance and combination of the acoustic cues to the prediction of voicing depend upon the clarity of the word.

Conclusions

This investigation provided quantitative data regarding four acoustic parameters associated with the production of the voicing contrast of word-initial alveolar plosives by English-speaking children and adults, and a limited amount of data about how these values influence sophisticated adult English-speaking listeners' perceptions of this contrast.

Only one articulatory context in which voicing contrasts occur was studied; therefore, interpretations of the results are limited. Generally, the statistical results supported the investigator's hypotheses. Listeners' abilities to accurately perceive word-initial alveolar plosives improved as speaker age increased. Among speakers, VOT, F1 frequency at the cnset of the postconsonantal vowel, and plosive burst amplitude were found to differ significantly between voiced and voiceless plosives, and variability of these values generally decreased with age. As their ages increased, speakers generally became more consistent in their productions of stable acoustical data that appeared to enable more accurate listeners' perceptions of the voicing contrast.

The results of this investigation appeared to support the argument for VOT as a primary cue to the perception of voicing of perceptually distinct word-initial alveolar plosives. Spectral information at the onset of the postconsonantal vowel, and burst amplitude to a lesser extent, provided secondary information to aid the listeners' perceptions of these tokens. However, the impact of speaker age or perceptual ambiguity upon the prediction of the voicing contrast of word-initial alveolar plosives requires more investigation.

References

Abramson. A. S., & Lisker, L. (1968). Voice timing: Cross-language experiments in identification and discrimination. Haskins Laboratories Status Report of Speech Research, SR 13/14, 49-63.

Baken, R. J. (1987). Clinical measurement of speech and voice. Boston: College-Hill Press.

Barton, D., & Macken, M. A. (1980). An instrumental analysis of the voicing contrast in word-initial stops in the speech of four-year-old English-speaking children. Language and Speech, 23, 159-169.

Borden, G. J., & Harris, K. S. (1980). Speech science primer: Physiology, acoustics, and perception of speech. Baltimore: Williams & Wilkins.

Bruning, J. L. & Kintz, B. L. (1987). Computational handbook of statistics (3rd ed.). Glenview, IL: Scott, Foresman and Company.

Catts, H. W., & Jensen, P. J. (1983). Speech timing of phonologically disordered children: Voicing contrast of initial and final stop consonants. Journal of Speech and Hearing Research, 26, 501-510.

Cohen, J., & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. New York: Erlbaum.

Drew, C. J. & Hardman, M. L. (1985). Designing and conducting behavioral research. New York: Pergamon Press.

Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. Acta Oto-laryngologica, suppl. 257, 5-43.

Enochson, L. (1986). Accurate digital signal processing mandates antialiasing filters. Personal Engineering & Instrumentation News, August, 39-43.

Fant, G. (1973). Speech sounds and features. Cambridge, MA: MIT Press.

Forrest, K., & Rockman, B. K. (1988). Acoustic and perceptual analysis of word-initial stop consonants in phonologically disordered children. Journal of Speech and Hearing Research, 31, 449-459.

Gilbert, J. (1977). A voice onset time analysis of apical stop production in 3-year-olds. Journal of Child Language, 4, 103-110.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. MIT Research Laboratories Electronic, 101, 198-213.

Hodge, M. M. (1989). A comparison of spectral-temporal measures across speaker age: Implications for an acoustic characterization of speech maturation. Doctoral dissertation, University of Wisconsin - Madison, Wisconsin.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. Journal of the Acoustical

Society of America, 35, 84-92.

Huck, S. W., Cormier, W. H., & Bounds, W. G. Jr. (1974). Reading statistics and research. New York: Harper & Row.

Jamieson, D. G. (1989). CSRE 3.0: The Canadian speech research environment [Computer program]. London, Ontario: Speech Communication Laboratory, Department of Communicative Disorders, University of Western Ontario.

Keller, E. (1992). Signalyze [Computer program]. Seattle, WA: InfoSignal.

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. Journal of Speech and Hearing Research, 19, 421-447.

Kent, R. D. & Forner, L. L. (1979). Developmental study of vowel formant frequencies in an imitation task. Journal of the Acoustical Society of America, 65, 208-217.

Kent, R. D. & Read, C. (1992). The acoustic analysis of speech. San Diego, CA: Singular Publishing Group.

Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. The Journal of the Acoustical Society of America, 72, 379-389.

Kewley-Port, D., & Preston, M. (1974). Early apical stop productions: A voice onset time analysis. Journal of Phonetics, 2, 195-210.

Kirk, R. E. (1982). Experimental design: Procedures for the behavioral sciences. Monterey: Brooks/Cole Publishing.

Klatt, D. H. (1975). Voice onset time, frication and aspiration in word-initial consonant clusters. Journal of Speech and Hearing Research, 18, 686-706.

Klecka, W. R. (1980). Discriminant analysis. In Lewis-Beck, M. S. (Ed.), Series: Quantitative applications in the social sciences. Newbury Park, CA: Sage Publications.

Kraemer, H.C. & Thiemann, S. (1987). How many subjects? Newbury Park, CA: Sage Publications.

Lachenbruch, P. A. (1975). Discriminant analysis. New York: Hafner Press.

Ladefoged, P. (1957). Three areas of experimental phonetics. London: Oxford University Press.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. Journal of Experimental Psychology, 52, 127-137.

Lindblom, B. E. F. (1963). Spectrographic study of vowel reduction. Journal of the Acoustical Society of America, 35, 1773-1781.

Lisker, L. (1975). Is it VOT or a first-formant transition detector? *The Journal of the Acoustical Society of America*, *57*, 1547-1551.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops. Acoustical measurements. *Word*, *20*, 384-422.

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

Macken, M. A. & Barton, D. (1979). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, *7*, 41-74.

Maxwell, E., & Weismer, G. (1982). The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice onset time for stops. *Applied Psycholinguistics*, *3*, 29-44.

McLaughlin, M. L. (1980). Discriminant analysis in communication research. In P. R. Monge & J. N. Capella (Eds.), *Multivariate techniques in human communication research* (pp. 175-202). New York: Academic Press.

Menyuk, P. & Klatt, M. (1974). Voice onset time in consonant cluster production by children and adults. *Journal of Child Language*, *2*, 223-231.

Milenkovic, P. (1989). CSpeech [Computer program]. Madison: Electrical and Computer Engineering, University of Wisconsin-Madison>

Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. Journal of the Acoustical Society of America, 85, 2114-2134.

Minifie, F. D., Hixon, T. J., & Williams, F. (Eds.). (1973). Normal aspects of speech, hearing, and language. Englewood Cliffs, NJ: Prentice-Hall.

Monsen, R. & Engebretson, A. (1983). The accuracy of formant frequency measurements. Journal of Speech and Hearing Research, 26, 89-97.

Moore, B. & Glasberg, B. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. Journal of the Acoustical Society of America, 74, 750-753.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. Journal of the Acoustical Society of America, 85, 2088-2113.

Nearey, T. M. (1992). Applications of generalized linear modelling to vowel data. In J. J. Ohala, T. M. Nearey, B. L. Derwing M. M. Hodge, & G. E. Wiebe (Eds.), The International Conference on Spoken Language Processing 92 Proceedings (Vols. 1-2, pp. 583-586). Edmonton, Alberta, Canada: University of Alberta

Ohala, J. (1972). How is pitch lowered? Journal of the Acoustical Society of America, 52, 124.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. Journal of the Acoustical Society of America, 75, 224-230.

Ohde, R. N. (1985). Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children. Journal of the Acoustical Society of America, 78, 1554-1561.

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

Revoile, S., Pickett, J. M., Holden-Pitt, L. D., Talkin, D., & Brandt, F. D. (1987). Burst and transition cues to voicing perception for spoken initial stops by impaired- and normal-hearing listeners. Journal of Speech and Hearing Research, 30, 3-12.

Shriberg, L., & Kent, R. (1982). Clinical phonetics. New York: Wiley.

Shriberg, L., & Kwiatkowski, J. (1980). Natural process analysis (NPA): A procedure for phonological analysis of continuous speech samples. New York: Wiley.

Steiner, V. G. & Pond, R. E. (1979). Preschool language scale. Columbus, OH: Charles E. Merrill Publishing.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. Journal of the Acoustical Society of America, 55, 653-659.

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. Journal of the Acoustical Society of America, 62, 435-448.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. Journal of the Acoustical Society of America, 79, 1086-1100.

Templin, M. C., & Darley, F. C. (1968). Templin-Darley tests of articulation. Iowa City: University of Iowa, Bureau of Educational Research and Services, Division of Continuing Education.

Umeda, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech. Journal of the Acoustical Society of America, 70, 350-355.

Weismer, G. (1981). Temporal characterisucs of the laryngeal devoicing gesture for voiceless consonants and fricative-stop clusters: Influences of vowel environment and speaker age. Journal of the Acoustical Society of America Suppl. 1 69, S68.

Wilson, F. B. (1971). The voice-disordered child: A descriptive approach. *Language, Speech, and Hearing Services in Schools*, *1(4)*, 14-22.

Zlatin, M. A., & Koenigsknecht, R. A. (1975). Development of the voicing contrast: Perception of stop consonants. *Journal of Speech and Hearing Research*, *18*, 541-553.

Zlatin, M. A., & Koenigsknecht, R. A. (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research*, 19, 93-111.

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, *68*, 1523-1525.

Appendix A

CALCULATION OF SAMPLE SIZE

(following procedure for multiple regression outlined by Cohen & Cohen, 1975)

At an alpha level of .05 and a study power of .80 (beta level of .20), given 4 independent variables, $L = 11.94$:

$$f^2 = \frac{R^2}{1 - R^2}$$

$$= \frac{.20}{.80}$$

$$= .25$$

where R^2 equals the multiple coefficient of determination, K equals the number of independent variables, L equals the value of Wilk's lamda, and N equals the number of subjects required:

$$N = \frac{L + K + 1}{f^2}$$

$$= \frac{11.94 + 4 + 1}{.25}$$

$$= 53 \text{ subjects required}$$

Appendix B

Summary of the number of potential subjects who did not pass the screening procedures according to speaker age.

	Speaker Age - Level			
Screening Procedure	**2.5 years**	**4.5 years**	**10 years**	**Adults**
Articulation	2	2		0
Language		1	0	0
Hearing		3	1	1
Voice	0	2	2	1
TOTAL	4	8	3	2

Appendix C

WORD LIST FOR PRODUCTION TASK IN ORDER OF PRESENTATION

1. *dot
2. caught
3. pop
4. got
5. *tot
6. bee
7. *tot
8. caught
9. *dot
10. toe
11. *tot
12. caught
13. *dot
14. key
15. coat
16. caught
17. *tot
18. pea
19. *dot
20. caught
21. got
22. *tot
23. got
24. *dot
25. coat
26. got
27. *tot
28. got
29. pot
30. caught
31. *dot
32. got

* indicates words used for perceptual and acoustical analyses.

Appendix D

FORMULA FOR BARK TRANSFORMATION

(Zwicker and Terhardt, 1980)

$$B=13 \arctan (0.76f) + 3.5 \arctan (f/7.5)$$

where B=critical band value in Bark

f=frequency in kHz

Appendix E

Group means (and standard deviations) for voice onset time (in ms) according to gender, speaker age, and voicing condition.

	Speaker Age - Level							
	2.5 years		4.5 years		10 years		Adults	
	M	F	M	F	M	F	M	F
/d/	22.32 (14.63)	13.11 (16.66)	18.57 (5.80)	18.53 (8.00)	21.54 (5.43)	24.49 (8.95)	-16.50 (39.37)	20.95 (5.87)
/t/	79.11 (37.50)	86.56 (31.58)	92.08 (29.06)	93.02 (20.69)	81.54 (9.23)	87.44 (17.97)	81.64 (18.50)	93.61 (11.21)

Appendix F

Group means (and standard deviations) for the first formant (F1) at onset of the postconsonantal vowel /ɑ/ according to gender, speaker age, and voicing condition.

	Speaker Age - Level							
	2.5 years		4.5 years		10 years		Adults	
	M	F	M	F	M	F	M	F
				Values in Hertz				
/d/	825.46 (97.95)	972.73 (202.19)	754.13 (62.40)	768.40 (74.86)	643.60 (104.13)	659.66 (72.56)	504.47 45.01)	624.38 (59.96)
/t/	1127.84 (174.08)	1352.27 (159.31)	1146.69 (101.69)	1222.27 (147.93)	960.51 (146.21)	953.73 (108.18)	634.53 (27.52)	922.69 (91.49)
				Values in Bark				
/d/	7.26 (0.67)	8.18 (1.10)	6.78 (0.47)	6.88 (0.54)	5.85 (0.75)	6.05 (0.56)	4.79 (0.40)	5.77 (0.49)
/t/	9.10 (1.01)	10.19 (0.73)	9.29 (0.57)	9.66 (0.81)	8.15 (0.96)	8.13 (0.68)	5.86 (0.23)	8.01 (0.66)
				Values in Log				
/d/	6.69 (0.11)	6.85 (0.17)	6.62 (0.08)	6.63 (0.10)	6.44 (0.14)	6.48 (0.10)	6.21 (0.08)	6.42 (0.09)
/t/	6.99 (0.16)	7.14 (0.15)	7.02 (0.09)	7.06 (0.16)	6.84 (0.16)	6.84 (0.11)	6.45 (0.05)	6.82 (0.10)

Appendix G

Group means (and standard deviations) for fundamental frequency (FO) at onset of the postconsonantal vowel /ɑ/ according to gender, speaker age, and voicing condition.

	Speaker Age - Level							
	2.5 years		4.5 years		10 years		Adults	
	M	F	M	F	M	F	M	F
				Values in Hertz				
/d/	298.36 (38.37)	340.43 (32.88)	290.77 (32.74)	298.42 (31.28)	246.61 (33.42)	223.82 (22.12)	106.68 (17.51)	188.49 (15.59)
/t/	306.98 (45.22)	320.45 (42.32)	294.73 (25.67)	305.11 (31.68)	248.63 (30.28)	246.94 (20.96)	115.07 (12.34)	211.78 (19.68)
				Values in Bark				
/d/	2.90 (0.36)	3.29 (0.30)	2.83 (0.31)	2.90 (0.29)	2.41 (0.32)	2.20 (0.23)	1.05 (0.17)	1.85 (0.15)
/t/	2.97 (0.42)	3.11 (0.40)	2.87 (0.24)	2.96 (0.30)	2.43 (0.29)	2.42 (0.20)	1.13 (0.12)	2.08 (0.19)
				Values in Log				
/d/	5.69 (0.12)	5.81 (0.09)	5.66 (0.11)	5.69 (0.11)	5.50 (0.15)	5.41 (0.11)	4.66 (0.16)	5.23 (0.09)
/t/	5.71 (0.13)	5.76 (0.14)	5.68 (0.09)	5.71 (0.11)	5.49 (0.18)	5.50 (0.09)	4.47 (0.10)	5.35 (0.09)

Appendix II

Group means (and standard deviations) for ratios of plosive burst amplitude relative to vowel amplitude (in Volts) according to gender, speaker age, and voicing condition.

	Speaker Age - Level							
	2.5 years		4.5 years		10 years		Adults	
	M	F	M	F	M	F	M	F
/d/	1.10	1.02	0.80	0.83	1.06	1.02	0.77	1.03
	(0.39)	(0.37)	(0.22)	(0.27)	(0.32)	(0.39)	(0.37)	(0.24)
/t/	1.71	1.61	1.53	1.36	1.24	1.42	1.07	1.27
	(0.58)	(0.75)	(0.62)	(0.48)	(0.29)	(0.59)	(0.59)	(0.48)

Appendix I

CALCULATION OF EFFECT SIZES FOR THE PERCEPTUAL AN YSES

(based on calculation of effect sizes for paired c .sons described by Kraemer & Thiemann, 1987)

$$\Delta = \quad / \; (\; ^{2} + 1 / pq)^{1/2}$$

$$\delta = (\mu_x - \mu_y) \; /$$

where Δ equals the effect size, δ equals Glass's effect size, μ equals the group mean, equals the p ed variance between the two groups, p equals the proportion of subjects within the first group to the total subjects in both groups, ar q equals the proportion of subjects within the second group to the total subjects in both groups.

Effect sizes were calculated for the largest and smallest significant differences among the speaker groups obtained for correct, incorrect and ambiguous judgments. Results were as follows:

Comparison	Effect Size	Category
	Correct Judgments	
2.5 yrs and adults	0.76	large effect
2.5 yrs and 10 yrs	0.41	medium effect
	Incorrect Judgments	
2.5 yrs and adults	0.26	small effect
2.5 yrs and 4.5 yrs	0.36	small effect
	Ambiguous Judgments	
2.5 yrs and adults	0.59	medium effect
2.5 yrs and 4.5 yrs	0.30	small effect