INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.
2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.
3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*
4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

***For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.**

UMI University Microfilms International

8601681

Nittrouer, Susan

THE ROLE OF COARTICULATION IN THE PERCEPTION AND PRODUCTION OF SPEECH BY YOUNG CHILDREN (3 TO 7 YEARS)

City University of New York PH.D. 1985

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

Copyright 1985
by
Nittrouer, Susan
All Rights Reserved

THE ROLE OF COARTICULATION IN THE PERCEPTION AND PRODUCTION OF SPEECH BY YOUNG CHILDREN

(3 TO 7 YEARS)

Susan Nittrouer

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1985

COPYRIGHT BY
SUSAN NITTROUER
1985

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

July 23rd 1985
date

Michael Studdert-Kennedy
Michael Studdert-Kennedy, Ph.D.
Chairman of Examining Committee

July 23rd 1985
date

Arthur Boothroyd, Ph.D.
Executive Officer

Arthur Boothroyd, Ph.D.

Lawrence Raphael
Lawrence Raphael, Ph.D.

Bruno Repp, Ph.D.

The City University of New York

Abstract

THE ROLE OF COARTICULATION IN THE PERCEPTION AND PRODUCTION OF SPEECH BY YOUNG CHILDREN

(3 TO 7 YEARS)

by

Susan Nittrouer

Advisor: Michael Studdert-Kennedy

Simultaneous movement of the articulators used to produce neighboring segments (or, coarticulation) is ubiquitous in speech. Its consequences can be seen in the acoustic structure of the signal, and perceptual experiments have demonstrated that adult listeners are sensitive to these consequences. However, whether coarticulation is a mandatory property of speech production, and use of its consequences mandatory in perception, remains debatable. The purpose of this study was to help define the role of coarticulation in the perception and production of speech.

Forty subjects (eight children at each of the ages 3, 4, 5, and 7 years, and eight adults) identified tokens from a synthetic /ʃ/-/s/ continuum followed by one of four natural vocalic portions: /i/ or /u/, spoken with transitions appropriate for either /ʃ/ or /s/. Children demonstrated larger shifts in fricative phoneme boundaries as a function of vocalic transition than adults, but relatively smaller shifts as a function of vowel quality; responses were less consistent for children than for adults; and, differences between children and adults decreased as children increased in age.

In a second experiment, samples were obtained from the eight adults and from two children in each age group uttering the syllables /ʃi/, /si/, /ʃu/, and /su/. Both the center of gravity of the fricative spectrum and second formant values, at several sample points, were measured. Results for all children, regardless of age, were similar with respect to adult utterances: vowel context more strongly affected their fricative spectra, but vocalic formant transitions were reduced. These results suggest that, in fricative-vowel sequences, children tend to coarticulate (or coproduce) the two segments more strongly than adults.

Overall, these results indicate that coarticulatory effects in production, and sensitivity to these effects in perception, are present at as young as 3 years of age. This finding was interpreted as suggesting that coarticulation is a madatory property of speech production, and that its acoustic consequences contribute to speech perception even in a child at the early stages of learning to speak. Developmental trends suggested that the organization of speech gradually becomes more segmental in nature, with changes occurring earlier in perception than in production.

This work is dedicated to every deaf child with whom I have had the opportunity to work. It is because of these children that I began to wonder how it is that anyone ever learns to speak and understand speech.

ACKNOWLEDGEMENTS

The completion of a Ph.D. thesis can be an emotional experience for some candidates. It certainly has been for me. What began as a dream some seven or eight years ago is becoming a reality. I met many people along the way who helped me to grow academically, and encouraged me to continue when I felt I would never reach the end. I am grateful to each of these individuals.

Specifically regarding my thesis, I want to express appreciation to my major advisor, Michael Studdert-Kennedy, for devoting so much of his precious time to this endeavor. Michael shared his ideas with me, listened to mine, and with all of this, helped me to develop into an independent researcher. The deaf children with whom I worked caused me to start wondering how it is that anyone learns to speak and understand speech; Michael helped me discover ways of framing the questions so that we may start finding some answers.

I am also grateful to the members of my supervisory committee. Arthur Boothroyd has been a friend and teacher for more years than either of us cares to see printed. He was there in the beginning to encourage me to pursue my Ph.D., and was here in the end to serve on this committee.

Larry Raphael offered many valuable suggestions which improved the quality of this work, and his good-natured flexibility made the doing of the work much easier.

My outside examiner, Bruno Repp, played a larger role in this project than even he may realize. In addition to his direct input to the final draft, many of his studies have served as models for me of how research should be done. So, he indirectly affected this study at every stage of its development.

None of the work reported here would have been possible without Haskins Laboratories. A grant to this facility (NICHD Grant HD-01994) provided part of the funding for this work; people at this facility provided a great deal of advice and encouragement for me. In particular, I thank Doug Whalen for his help with programming, suggestions regarding method, and comments on an earlier draft. I also thank Katherine Harris for serving as a reader.

Several people helped with the actual testing. They were Denise Wright, Melanie Campbell, and Lise Jensen. I am grateful to all of them for their contributions.

Finally, I want to thank all the children who participated in this study for patiently playing my silly games, and their mothers for welcoming me into their homes. Especially, I would like to thank my nephew and niece, Todd and Julie Selvaggi, and my young friend, Albert Rothstein. Many pilot studies preceded the experiments reported here, and these three children were often first to try new procedures. The efforts of my sisters, Karen Selvaggi and Patricia Nittrouer, in enlisting subjects are also gratefully acknowledged.

TABLE OF CONTENTS

List of Tables

Table page

Perception Experiment

Production Experiment

List of Figures

Figure page

Perception Experiment

Production Experiment

List of Appendices

Appendix page

Introduction

When two people communicate through spoken language, the ideas of one person are transmitted to the other by a complex series of events. The speaker's ideas are structured into sentences and words, which in turn consist of smaller linguistic units (morphemes, syllables, phonemes), and these are transformed into acoustic signals. The listener receives the acoustic signals, recovers the linguistic units, and ultimately the original ideas. Before the 1940's, these linguistic units were largely studied without the aid of instruments to analyze the acoustic signal. Phoneme-sized phonetic segments (the postulated basic units of linguistic communication) were described in the terminology of traditional articulatory phonetics.

With the development of the sound spectrograph, linguists began to seek the acoustic correlates of phonetic categories. This task proved to be no simple matter. Joos (1948) described the two major difficulties. First, lines, vertical to the time axis, could not be clearly drawn between individual segments. While the center of a segment might be discernible in the acoustic pattern, regions existed between segments that did not seem to belong uniquely to one or the other. Joos termed this difficulty the problem of segmentation (cf. Fant, 1962). The second difficulty was that the acoustic attributes of both the center and adjoining regions of a given segment varied depending on the particular phonetic context. This finding has come to be known as the problem of invariance (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Pisoni, 1985). While lack of segmentation and lack of invariance are often described as separate problems, they actually result from the same cause: simultaneous movement of the articulators used to produce different neighboring segments, a phenomenon known as coarticulation.

Although the effects of coarticulation are clearly observable in speech spectrograms, many linguists and psychologists have continued to believe that invariant acoustic cues for phonetic categories exist and can be discovered if the signal is only analyzed appropriately. In the early 1950's the goal was to compile a "code book ... [with] one column for acoustic entries and another column for message units" (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952, p. 605). To this end, perceptual studies were conducted with synthetic stimuli, such that all aspects of the signal were held constant, except one. This component was systematically varied, and the effects on perception measured. A principal outcome of this work was to show that listeners respond differently to the same acoustic component, and similarly to different components, depending on the phonetic context -- very much as might have been predicted from spectrographic studies. For example, Liberman, Delattre, and Cooper (1952) reported that identical schematic noise bursts, synthesized on the Pattern Playback (Cooper, Liberman, & Borst, 1951), led to the perception of either /p/ or /k/ depending on the following vowel. Similarly, Cooper, Delattre, Liberman, Borst, and Gerstman (1952) found that second-formant (F2) transitions with identical trajectories cued different stops depending on the following vowel quality: Falling transitions cued velar stops when followed by front vowels but alveolar stops when followed by back vowels. Conversely, flat transitions cued alveolar stops when placed before front vowels but velar stops when placed before back vowels.

While perceptual research was going on with synthetic speech, work in the analysis of natural speech was demonstrating that the patterns of perceptual results had their bases in production. For example, it was found that formant transitions for consonants differ as a function of vowel environment (e.g., Öhman, 1966). Further, Lindblom (1963) demonstrated that vowels, often considered to be more invariant than consonants, are not exempt from contextual variation: formant frequencies, even during the vocalic steady-state, vary as a function of consonantal environment.

One result of these studies of natural and synthetic speech was that the concept of the phoneme-sized phonetic segment as the basic unit of perception was called into question, and the possible role of the syllable began to be considered. Therefore, interest developed in describing the structure of the syllable. This interest, in turn, emphasized the importance of defining the roles of coarticulation in production and perception. Specifically the questions were: Is coarticulation essential to production, and is it necessary to take into account the acoustic consequences of this coarticulation in speech perception? Or, instead, is coarticulation a learned behavior resulting from a speaker becoming rapid and skillful? If so, then are the acoustic consequences of coarticulation merely a source of noise, to be filtered out?

These two contrasting views of coarticulation might be termed the code theory and the cipher theory. The distinction between code and cipher was first drawn by Liberman et al. (1967). The acoustic structure would be a cipher on the linguistic message if each acoustic segment corresponded to a single linguistic segment, and vice versa. However, according to Liberman et al. (1967), the acoustic structure is a code on the linguistic message: The relation is both many-to-one and one-to-many, rather than one-to-one. Thus according to the code theory, phonemes are restructured at the acoustic level to permit rapid transmission. The syllable can then be viewed as a package containing the restructured phonemes. The listener was said to recover the phonetic string because he is also a speaker, and so knows how the restructuring occurs.

From this perspective, coarticulatory effects are an essential aspect of production, the acoustic results of which must be used for perception. Specifically, adjacent segments overlap in production in order that speech may be transmitted at rapid rates. Therefore, the acoustic characteristics of phonetic segments vary as a function of the phonetic contexts in which they occur. When one listens to speech, one somehow "adjusts" the acoustic values that will be accepted as characteristic of a given segment based on the phonetic environment in which it occurs, and thus recovers the phonetic segment originally encoded by the speaker.

An alternative approach to defining the role of coarticulatory effects is a cipher theory, such as that offered by the theory of acoustic invariance. This is the position proposed by Stevens and Blumstein, and described in a number of papers (e.g., Stevens, 1975; Stevens & Blumstein, 1978; Blumstein & Stevens, 1981). Essentially this theory contends that the syllable, if properly described, exhibits regions of static acoustic information that remain constant across phonetic contexts. These static regions are present in the acoustic signal because articulatory targets are generally regions of high acoustic stability (Stevens, 1972). That is, even though coarticulation may cause the exact articulatory position for any phoneme to vary across phonetic contexts, the acoustic results for such variants of a given target will be similar. The resulting static acoustic information is termed a 'property', and property-detecting mechanisms are postulated to exist in the auditory system.

Failure to discover these invariant acoustic properties in earlier work, according to these authors, was the result of improper analysis procedures. Most earlier work had assumed that perception was based on dissection of the spectral structure of the signal into its component formants --- the changing frequencies of the spectral peaks. Stevens and Blumstein, by contrast, suggest that perception is based on the whole spectrum, integrated over some appropriate time window, which takes into account additional acoustic features such as relative formant amplitudes and bandwidths, but neglects changes over time. According to this view, then, contextual effects are simply the result of the articulators moving from one static region (segment) to another and have no true bearing on perception. Infants are hypothesized to be born with the property detectors necessary to identify segments based on these static characteristics (Stevens & Blumstein, 1978, p. 1367). Then, because certain transitional patterns are consistently associated with specific invariant properties, the child learns to use these transitional patterns for perception when the acoustic regions exhibiting the invariant properties are absent or, as in synthesis studies, set at ambiguous values. Thus, in the Cooper et al. study (1952), subjects were able to use the context-variant formant transitions to identify stops when the bursts were absent. The burst, then, can be seen as the 'primary' cue, and the formant transition as the 'secondary' (or learned) cue.

These two theories differ in their views of what the listener does with the acoustic consequences of coarticulation. Whalen (1984) has labeled these alternative views the "integrating" and "disposing" accounts. Supporters of the speech as a code viewpoint suggest that the listener uses all the information in the acoustic signal for perception of a phoneme, including information resulting from coarticulation (the integrating account). According to Whalen, proponents of the acoustic invariance theory believe that only the invariant information corresponding to phonemic categories is used for perception, and information resulting from coarticulation is normally discarded (the disposing account). Where these two theories agree is in their assumption that the final percept, whether the output of an acoustic property detector or of a specialized decoding process, is a static phonological entity.

A third approach to understanding the role of coarticulation views the linguistic percept as intrinsically dynamic (Fowler, Rubin, Remez, & Turvey, 1980). This approach does not support a view of the percept as an underlying discrete, static segmental unit; rather it suggests that the articulatory movement itself or its underlying control structure, is the percept. According to this account, speakers do not plan to produce invariant articulatory positions (which are then restructured to fit their context), but instead produce a sequence of overlapping movements, which characterize segments. (Fowler et al. use the term "coordinative structures" to emphasize that the several articulators are not just moving simultaneously and separately; instead groups of muscles function as units to perform designated patterns of movement.) The overlapping movements associated with neighboring segments sum to produce a complex movement vector -- and, of course, a correspondingly complex acoustic pattern. Fowler (1980) introduced the term "coproduction" (rather than coarticulation) to emphasize that the movement vectors are not the result of adjusting one segment to the context of another, but are simply the summed output of independent (and invariant) motor control structures corresponding to segments. According to this account, the vectors can be resolved and individual segments recovered from the acoustic pattern because they remain internally coherent and qualitatively distinct.

An analogy given by Fowler et al. is that one is able to hear the voice quality of a singer as distinct from accompanying instrumental music, even though the waveforms from both sources combine, because components of the waveform arise from different sources with different dynamic characteristics. Each sound is produced in a different way and results in qualitatively different acoustic patterns. Similarly, vowels and consonants are produced in different manners. Vowels are produced by relatively slow, global changes in the vocal tract. Consonants, on the other hand, are produced by rapid, local obstructions. Consequently, vowels result in slowly changing patterns of formant frequencies, and consonants result in rapid changes superimposed on this slower background pattern. Thus, the listener is perceiving continuous, overlapping patterns of movement that correspond to the traditional linguistic entities of phonological description (i.e. phonemes). As with the code theory, the dynamic theory of coarticulation suggests that listeners use all available acoustic information for perceiving speech (the integrating account of acoustic cues).

Three different theories, then, have been suggested to account for coarticulation in production, and sensitivity to coarticulatory effects in perception: the code theory, the theory of acoustic invariance, and the dynamic theory. It is difficult to determine which of the three theories most accurately describes the role of coarticulation in speech production and perception, a goal which would have a variety of benefits. For example, Levinson and Liberman (1981) suggest that significant improvements in automatic speech recognition may not be achieved until a better understanding of human speech perception is available.

One problem in deciding among the three possible accounts of coarticulatory effects offered here is that these theories are primarily based on results from studies of adult speech. Adults clearly demonstrate coarticulatory effects in production and sensitivity to these effects in perception, but whether all these effects are necessary components of the processes or the results of becoming highly efficient communicators (in which case these effects would not be necessary components) cannot be determined from these studies.

One possible way of assessing the function of coarticulation in adults' speech would be to investigate coarticulatory effects in children's speech. This approach views the adult's speech system as developing naturally from the child's. Therefore, by studying children's speech production and perception and how they develop, a better understanding might be gained of adults' speech. Several results are possible from this sort of investigation, and are illustrated in Table 1.

Interpretation of possibilities #1 and #4 would be relatively straightforward. If no evidence were found of coarticulatory effects in children's production, or sensitivity to these effects in their perception (possiblity #1), this would lend support for the idea that coarticulation is not essential to production, and use of its consequences is not necessary for perception. If evidence of coarticulation were found in children's production, and sensitivity to its acoustic consequences in perception (possibility #4), this would suggest that coarticulation may play a necessary role in the speech system. The interpretation of possibilities #2 or #3 would be more tentative. If it were found that children demonstrate sensitivity to coarticulatory effects in perception, but do not coarticulate in production (possibility #2), this might indicate that the use of information about coarticulation is a necessary component of speech perception, but experience is required before coarticulation occurs in production. If it were found that children demonstrate coarticulatory effects in production, but do not use information resulting from

Table 1: Possible results for the present study

Do children display evidence of coarticulation, or sensitivity to its acoustic consequences, in:

perception?		no	yes
production?	no	#1 perc.=no prod.=no	#2 perc.=yes prod.=no
	yes	#3 perc.=no prod.=yes	#4 perc.=yes prod.=yes

coarticulation in perception (possibility #3), then we might conclude that coarticulation is simply a natural consequence of speech production, the acoustic results of which need not be used in perception.

One of the few articles to address the question of children's coarticulation confines its attention to production. Kent (1983) suggests that certain coarticulatory effects are less evident in the speech of 4-year-old children than of adults, perhaps because they are less efficient speakers, producing one phoneme at a time. Other researchers (e.g., Menyuk & Menn, 1979), however, argue that younger children (roughly 1 and 2-year-olds, during the early stages of vocabulary growth) produce speech as complex "prosodic" units equivalent to syllables or words and develop more analytic linguistic strategies only later. If this pattern were continued into later development, we would expect that coarticulatory effects might be even more evident in children's speech production than in adults', which would contradict Kent's description.

The purpose of the present research was to investigate a particular class of coarticulatory effects in both perception and production of speech by young children. First, the question of the extent to which children use these coarticulatory effects for the perception of speech will be addressed. Then, the role of coarticulation in production will be investigated in a study to be described separately. Finally, the results of both the perception and production experiment will be discussed in terms of what they might have to say about coarticulation in the adult's speech system.

Experiment 1: Perception

REVIEW OF THE LITERATURE

Researchers interested in listeners' abilities to use the acoustic consequences of coarticulation in speech perception have primarily conducted experiments varying the phonetic context in which segments are presented. The assumption underlying these experiments is that variation in perceptual results as a function of phonetic context reflects listeners' abilities to use their knowledge of coarticulation and its acoustic consequences for perception. Although several contextual effects have been investigated, the present experiment focused on just one: the influence of vocalic segments on fricative identification.

Adult Studies

Contextual effects in adults' speech perception have been studied most frequently by presenting tokens from the same acoustic continuum in different phonetic contexts. For example, pole (i.e., formant) frequencies of stimuli along an /ʃ/-/s/ fricative continuum might vary in steps of 100 Hz from the most /ʃ/-like stimulus to the most /s/-like. These stimuli would be followed by vocalic segments differing in some way, such as vowel quality. Subjects identify the combined tokens as beginning with either /ʃ/ or /s/, and phoneme boundaries are determined. A phoneme boundary (also known as a "crossover point") is the point on the acoustic continuum at which a listener changes from hearing mostly segments belonging to one phonemic category to hearing mostly segments belonging to the other category. Thus, the phoneme boundary is the point at which the probability of either response is .50. If the placement of this boundary is shifted as a function of the phonetic environment, then it is assumed that perception of the segment given by the acoustic continuum was affected by the segments with which it was combined.

Kunisaki and Fujisaki (1977) used this method to investigate vowel context effects in fricative perception. They prepared a ten-step synthetic fricative continuum (from /ʃ/ to /s/), followed by synthetic /a/ and /u/ vocalic portions. These vocalic portions contained first and second formant transitions, but it is unclear from the description whether the transitions were appropriate for /ʃ/ or /s/. Nonetheless, these authors demonstrated that listeners perceived more stimuli as /s/ when the fricative noise was followed by the rounded vowel /u/ than when it was followed by the unrounded vowel /a/. The explanation proposed by Kunisaki and Fujisaki was that, in natural speech, anticipatory liprounding lowers the pole values of the fricative, so that listeners come to expect that /s/ will have lower pole values when the following vowel is rounded than when it is unrounded. In other words, listeners know the acoustic results of liprounding and use this knowledge in their judgements of a synthetic series. The contextual effect in perception is thus linked to coarticulation in production.

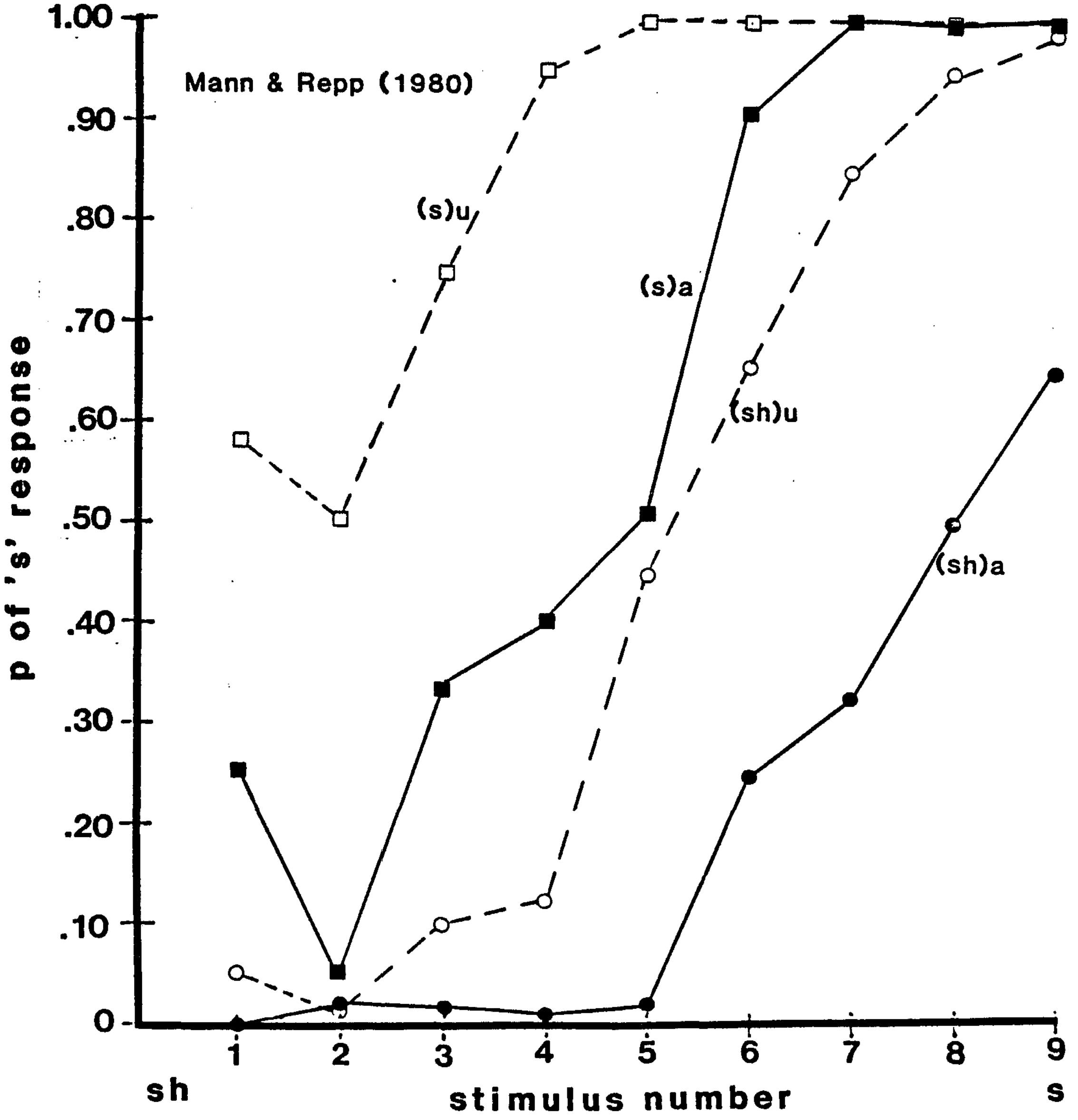
An alternative explanation for this effect might be based strictly on properties of general audition. For example, fricative noises with lower frequency values may be perceived as relatively higher (i.e., more /s/-like) in the /u/ context due to auditory contrast induced by the lower spectrum for /u/ than for /a/.

Mann and Repp (1980) replicated the vowel context effect of Kunisaki and Fujisaki (1977) using a nine-step synthetic /ʃ/-/s/ continuum and synthetic vowels corresponding to /a/ and /u/ with formant transitions appropriate for an alveolar place of articulation, that is, for /s/. They also demonstrated an effect of the initial vocalic transition using these same fricative noises and natural vocalic segments of /a/ and /u/ with transitions appropriate for both /ʃ/ and /s/. Identification functions drawn from data obtained in this experiment are shown in Figure 1. It can be seen that shifts in phoneme boundaries were obtained as functions of both the vowel context and formant transition.

However, in another experiment with flat vocalic formants following the fricative noise, the vowel context effect virtually disappeared. Thus the question arose as to the actual source of the vowel context effect. Did the transition merely provide the coherence between fricative and vowel segments necessary for them to be heard as a single syllable to which both fricative and vowel contribute? Or was the vowel effect the result of the listener using information in the vocalic transitions rather than vowel quality _per se_? Or are transition and vowel context effects separate phenomena?

Evidence existed to suggest that the first possibility provided at least a partial explanation for why the vowel context effect disappeared when transitions were removed. Work by Darwin and Bethell-Fox (1977) and

Figure 1: Identification functions from Mann and Repp (1980) data, male speaker



by Dorman, Raphael, and Liberman (1979) had demonstrated the importance of signal continuity in providing phonetic coherence. For example, in the Darwin and Bethell-Fox experiment the perception of a liquid-vowel syllable was changed into the perception of a stop-vowel syllable when fundamental frequency shifted in the region of the segmental boundary. Moreover, Mann and Repp (1980) noted that, in the transitionless stimuli, the fricative and vowel segments sounded segregated.

The second possibility, that the transition was solely responsible for the context effect, was eliminated both by the first experiment of Mann and Repp (1980) and by Whalen (1981): both studies demonstrated a vowel context effect when all stimuli provided transitional information appropriate for the same fricative. This effect can be seen in Figure 1 from Mann and Repp.

Further, Whalen (1981) offered additional support for the third possibility, that both transitions and vowel context affect the placement of the phoneme boundary. Using synthetic fricative noises from an /ʃ/-/s/ continuum and vocalic segments appropriate for /i,u,ḯ,ǘ/ with both /ʃ/ and /s/ transitions, Whalen demonstrated the transition and the vowel context effects with both synthetic and naturally produced vowel segments. Thus any concern that these effects were artifacts of synthetic speech or of a quirk in the particular natural samples used could be dismissed. The fact that these effects were obtained for English speakers using non-English vowels (/ï/ and /ü/) suggests that the effects are not related to specific linguistic experience, but rather to some more basic characteristic of auditory or phonetic perception such as the effects of liprounding on fricative spectra.

Finally, the finding that vocalic transitions are required to obtain a vowel context effect emphasizes the fact that transition and vowel effects have the same roots. While the size of each effect may be measured separately, both effects seem to arise from the acoustic effects of coarticulation and from the listener's ability to use these acoustic effects in perception. Whether or not coarticulation is *essential* to production and use of its effects *essential* to perception remains unknown.

It seems plausible that some forms of coarticulation may be obligatory for production, and that the acoustic consequences of these forms may then be necessary for perception. Other forms of coarticulation may not be essential for production, and their acoustic consequences may be unnecessary for perception. For example, fricative-to-vowel transitions would seem to be an automatic consequence of producing a CV syllable with an initial fricative. These transitions would then provide the listener with direct information concerning fricative identity. Thus, it might be predicted that young children would display fricative-to-vowel transitions similar to adults in their productions, and would use information conveyed by these transitions for perception. On the other hand, liprounding during fricative production in anticipation of a rounded vowel is probably optional, and may reasonably result from a speaker becoming highly efficient. Therefore, children may not display this coarticulatory effect in either production or perception.

The integrating account of the perception of acoustic information (Whalen, 1984) suggests that the listener uses all available information, including that which results from coarticulation. The disposing account acknowledges that acoustic information is present in the signal as a result of coarticulation, but suggests that this information is unnecessary for identification of phonemic segments under normal conditions, and therefore, is ignored by the listener.

Child Studies

Only one study has investigated a phonetic context effect in children (Mann, Sharlin, & Dorman, 1983). In this study, the /ʃ/-/s/ continuum of Mann and Repp (1980) was combined with naturally produced vocalic segments taken from utterances of /ʃeIv/ and /ʃu/. Consequently, the vowel context effect could be studied, but not the transition effect. Subjects were adults, 5-year-olds, 7-year-olds with age-appropriate speech production skills, and 7-year-olds judged to misarticulate fricatives. In response to stimulus presentations, children were required to point to pictures illustrating the stimuli (e.g., a man shaving corresponded to the perception of /ʃeIv/). The resulting identification functions (pooled for each subject group) appear similar in one respect for the adults and all 3 groups of children: more 's' responses were given in the /u/ context. The sizes of the shifts in mean crossover points as a function of vowel context are also similar for all groups (a shift of approximately one stimulus on the continuum). Although values for slopes are not given, visual inspection of the identification functions suggests that they are equally steep for both groups of 7-year-olds and for adults, but are somewhat less steep for the 5-year-olds: in other words, the 5-year-olds were less consistent in their responses. If this is the case, this result would match findings of other studies. For example, Simon and Fourcin (1978) found that children between the ages of 2 and 14 years became more 'categorical' (that is, less variable) in their labeling of stimuli from voiced-voiceless initial stop continua with increasing age. Krause (1982) demonstrated that the slopes of identification functions for voiced-voiceless final stops were shallower for 3- and 6-year-old children than for adults, and that the functions were shallower for 3-year-olds than for 6-year-olds.

The Mann et al. (1983) study suggests that children as young as 5 years of age can use the coarticulatory effect of liprounding for fricative identification. It also shows that experience with one's own correct productions is not a necessary prerequisite for this perceptual skill because the misarticulating 7-year-olds performed similarly to the other 7-year-olds and adults. However, it should be noted that these children were judged to be *misarticulating* fricatives; no measurements were made of the extent to which they coarticulated fricatives with following vowels. Conceivably a child could coarticulate, while misarticulating. Therefore, we cannot rule out the possibility that a child needs experience listening to coarticulatory effects in his or her own speech before being able to use the acoustic consequences of these effects for perception. Moreover, this study does not rule out auditory contrast as an explanation for this vowel effect.

Additionally, the apparent differences in slopes between the 5-year-olds and the 7-year-olds and adults is of interest. The simplest explanation for these differences is that they reflect the variability usually associated with a partially learned (or not fully established) response. However, the degree of control exerted by formant transitions on responding was not assessed because formant transitions appropriate for /ʃ/ were used for all stimuli. If the responses of the 5-year-olds were more strongly controlled by formant transitions than the responses of 7-year olds and adults, the fixed transition stimuli of this study would probably make these stimuli particularly ambiguous for 5-year-olds. Thus, it seems of interest to investigate the transition effect in children's speech perception to see if they learn to use different kinds of coarticulatory information at different ages.

Although Parnell and Amerman (1978) did not specifically investigate phonetic context effects, they did look at the extent to which children are able to use information present in one segment to identify an adjacent segment. These researchers edited naturally produced samples of stop-vowel syllables into several smaller components:

burst + aspiration
burst + aspiration + vocalic transition
vocalic transition only
vocalic transition + vowel

Subjects were asked to identify the syllable from which the edited samples came. Consonant and vowel recognition scores were compared among 4-year-olds, 11-year-olds, and adults. Results indicated that 4-year-old children could not use the information present in these smaller sections to recognize phonemes as well as 11-year-old children and adults. For example, when presented with the burst + aspiration + vocalic transition, the 11-year-old children and adults correctly identified the vowel 85% and 83% of the time, respectively. The 4-year-old children correctly identified the vowel only 55% of the time. This result suggests that young children may not use information about coarticulation to as great an extent as adults. However, the experimental task required inference of a missing segment. Young children may have more difficulty with this task than adults.

One final study deserves mention. Morrongiello, Robson, Best, and Clifton (1984) investigated the abilities of 5-year-olds to use multiple cues in recognizing a stop between an initial consonant and following vowel. Briefly, natural /s/ segments were followed by synthetic /eI/ portions which were identical in all aspects except the onset values of F1 (430 Hz or 230 Hz). A silent interval varying in duration between 0 ms and 104 ms was inserted between these two segments. For adults, the crossover point between the identification of 'say' and 'stay' differs depending on whether the F1 onset supports the perception of 'say' (430 Hz) or 'stay' (230 Hz): relatively more 'stay' responses are given to tokens with the 230 Hz vocalic onset (Best, Morrongiello, & Robson, 1981). Morrongiello et al. found that, although children gave more 'stay' responses to tokens with the 230

Hz vocalic onset, the shift in phoneme boundaries as a function of onset was smaller than that demonstrated by adults. This decreased boundary shift exhibited by children was exclusively due to their giving more 'stay' responses to tokens with the 430 Hz vocalic onset. A possible explanation offered by Morrongiello et al. was that the children may have weighted the transitional information relatively more heavily, and the temporal information relatively less heavily, as compared to adults. That is, perhaps young children selectively attend to the transitional cues more than adults. This suggestion was based on the finding that even a small F1 transition seemed to be sufficient to support the perception of 'stay' for these children.

Present Study

The main purpose of the present experiment was to investigate the development of perceptual sensitivity to the acoustic consequences of coarticulation by determining whether children demonstrate the same sized shifts in phoneme boundaries as a function of phonetic context as adults. If the same sized shifts were seen, this would show that children use the acoustic consequences of coarticulation to the same extent as adults, implying that the use of this information occurs at an early age, and therefore, may be essential to speech perception. If children showed smaller shifts or no shifts, this would suggest that the ability to use information resulting from coarticulation is not essential to the perception of speech, but is an acquired skill -- perhaps of particular utility in following the "reduced" patterns of casual adult speech.

A second purpose of this experiment was to determine whether children demonstrate different sized boundary shifts depending on the kind of context effect involved. Several studies reviewed here investigated both the transition and vowel effects in adults' perception of fricatives. Conceivably, children could show the same results as adults for one of these effects, but different results for the other. This finding might imply that the abilities to use different kinds of coarticulatory information develop at different rates, possibly because some forms are mandatory in production while others are optional.

Finally, a third purpose of this research was to determine the extent to which children recover individual segments from the signal. If children do not recover individual segments to the same extent as adults, it might be predicted that responses would be less consistent for children than adults, indicating that relatively less weight is being assigned to the spectrum of the fricative noise *per se*, and more attention is given to the syllabic structure as represented by formant transitions.

For the present research, it was necessary to use stimuli that would permit observation of the effects of several different kinds of information. For this reason, hybrid syllables of synthetic fricative noises and natural vowels were prepared. The fricatives varied along a nine-step continuum from /ʃ/ to /s/, and the vocalic portions were excerpted from naturally spoken /ʃi/, /si/, /ʃu/, and /su/ syllables, so that formant transitions were appropriate for either /ʃ/ or /s/. Thus, the contributions of three sources of information could be studied: the spectrum of the fricative noise, the vowel context, and the formant transition. The contribution of the fricative noise spectrum would be reflected in the consistency of responses. The contributions of vowel context and formant transition would be reflected by the sizes of phoneme boundary shifts as functions of these two sources of information.

In addition, some support might be provided for either the integrating or disposing accounts of perception (Whalen, 1984). For example, if it were found that children do not demonstrate a transition effect, this would lend support to Stevens and Blumsteins' (1978) contention that the auditory system possesses property detectors which detect static, invariant characteristics of the acoustic signal, so that coarticulatory information is not required for speech perception (the disposing account). On the other hand, if children demonstrate both a vowel context and transition effect, then support would be provided for the notion that, even at an early age, all available acoustic information is considered in identifying phonetic segments (the integrating account).

METHOD

Subjects

One group of adults and 4 groups of children of the ages 3, 4, 5, and 7 years participated in this experiment. There were 8 subjects in each age group. All adults were 20- or 21-years-old, and all children were within -1 and +5 months of their designated age. No age group had greater than a 5 to 3 ratio between males and females. All subjects were native speakers of American English, speaking a dialect typical of the Middle Atlantic States, and all had age-appropriate articulation skills, as judged by two independent speech pathologists from recordings of spontaneous speech.

Each subject in this study also passed audiometric and tympanometric screenings. Pure tones of 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz were presented at 20 dB HL for the audiometric evaluation. Tympanometric results for all subjects displayed normal pressure peaks between +100 daPa and -150 daPa.

Finally, all subjects were required to identify accurately the endpoints of the /ʃ/-/s/ continuum followed by /i/ and /u/ with appropriate vocalic transitions at greater than 90% accuracy during an initial practice trial. In addition, any subject failing to show greater than 80% correct identification for these good exemplars during the actual testing was eliminated. This requirement served as a check on the attentiveness of the children.

It is worth noting that, with one exception, no child younger than 3 years 4 months was found who could meet these last 2 criteria. Several children between the ages of 3 years 0 months and and 3 years 2 months attempted the task, but failed either to meet the 90% criterion during practice, or the 80% criterion during actual testing. Only one child of 3 years 1 month met these 2 criteria; all other 3-year-old subjects were 3 years 4 months or 3 years 5 months. These children appeared to have no difficulty identifying the endpoints or attending to the task.

Stimuli

The stimuli were constructed on the basis of pilot tests, and consisted of tokens from an /ʃ/ to /s/ continuum created on the serial software synthesizer at Haskins Laboratories. Each token contained 210 ms of fricative noise with a single pole (spectral peak). There were 9 tokens on this continuum with pole values ranging from 2200 Hz to 3800 Hz in 200 Hz steps. Each token also had a zero at .75 x the pole value, which had the effect of creating a dip in the spectrum. The amplitude of these tokens rose gradually by 20 dB over the first 170 ms, remained constant for the next 20 ms, then fell by 5 dB over the final 20 ms. Fricative noises were synthesized with a sampling rate of 10 kHz, and were low-pass filtered at 4.9 kHz.

The vocalic portions of the stimuli were the vowels /i/ and /u/ produced by a male speaker in the syllables /ʃi/, /si/, /ʃu/, and /su/. The unrounded vowel /i/ was chosen instead of /a/ because second formant trajectories are more distinct between /i/ and /u/ than between /a/ and /u/ (Soli, 1981). The speaker originally produced 5 samples of each syllable, but because Mann and Repp (1980) had found little token-to-token variability in results, it seemed sufficient to use a single sample of each. Tokens to be used were selected so that they matched one another as closely as possible in duration and intonation contour. These syllables were digitized on a VAX computer using a 10 kHz sampling rate, and low-pass filtering with an upper cut-off of 4.9 kHz. The vowels, including the vocalic transitions, were isolated from the fricative noise using a waveform editing program. The vocalic segments were between 350 ms and 370 ms in duration. Spectrographic displays and LPC analysis were used to obtain frequency values for formant transitions. These values are given in Table 2 along with the transition values from Whalen (1981) for comparison, and the spectrograms are displayed in Figure 2. Thus, there were 9 fricative noises combined with 4 vocalic segments, resulting in a total of 36 unique stimuli. Each stimulus was presented to subjects 10 times, making a total of 360 test tokens.

Four test tapes were prepared: two using only the /i/ context, and two using only the /u/ context. Ten practice items began each tape (the 2 endpoints with appropriate vocalic segments presented 5 times each in random order). The 90 test items per tape were then presented in randomized blocks of 18. Before each group of 10 stimuli, a female voice announced the number of the next group of stimuli to be heard (numbers 1 through 9). The interstimulus interval was 2 s. Tapes were made on Scotch Audio Tape 208 at a tape speed of 7.5 inches per second and at a -5 dB recording level.

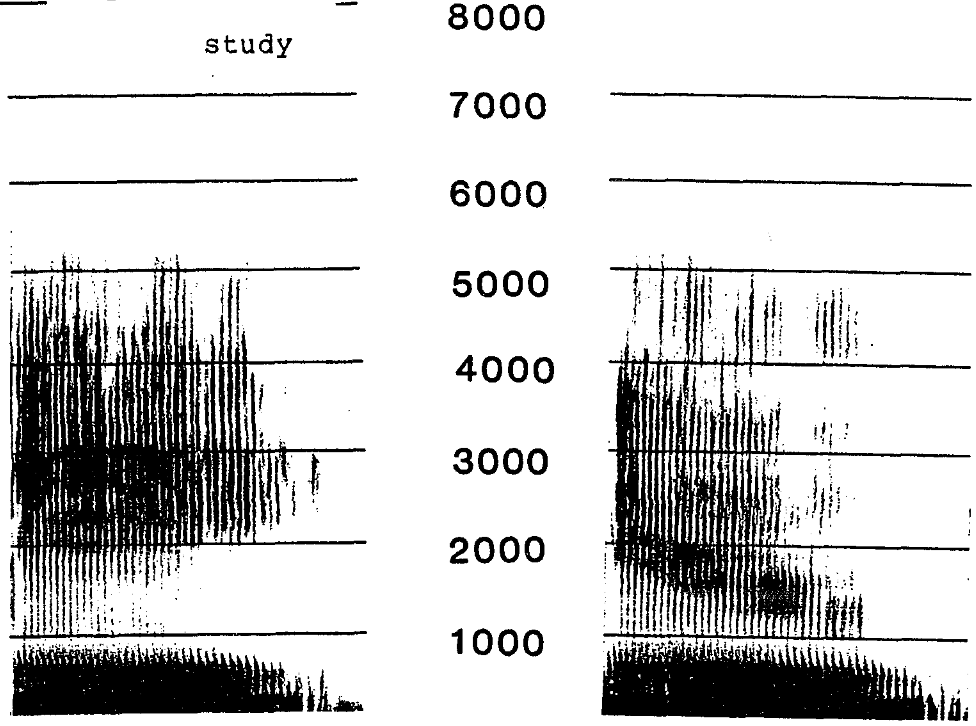
Materials and Equipment

Stimuli were presented via a Uher Model 4200 tape recorder using Sennheiser earphones. A group listening

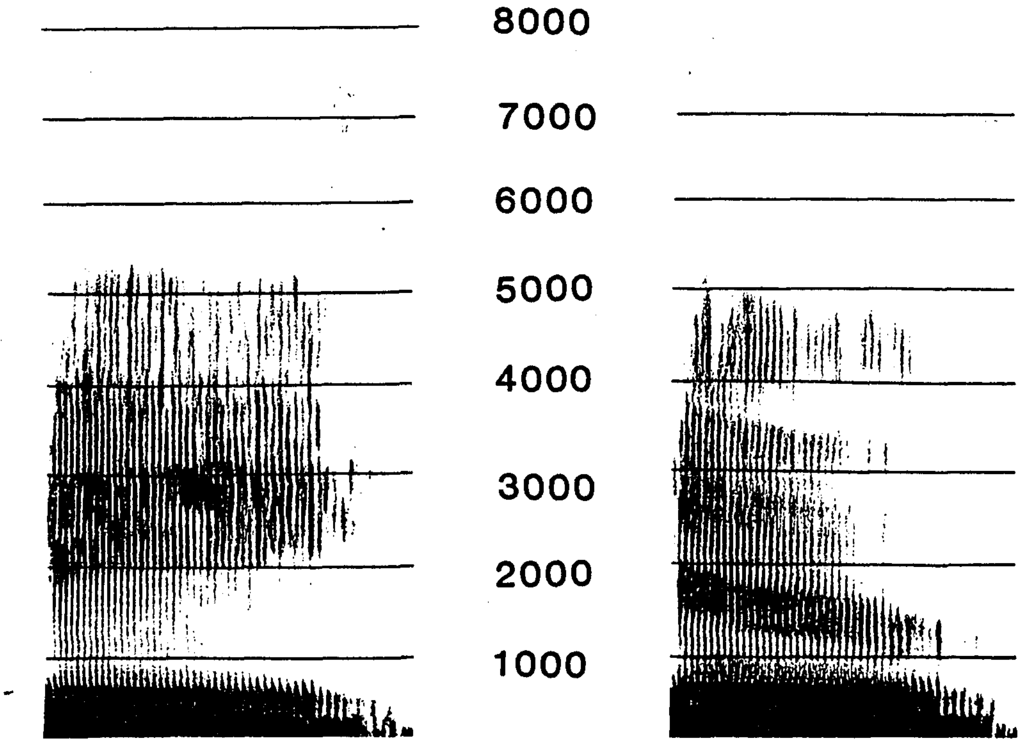
Table 2: Formant characteristics of vocalic portions used in this study and Whalen (1981)

	F2			F3		
	onset,	dur,	steady-state	onset,	dur,	steady-state
	(Hz)	(ms)	(Hz)	(Hz)	(ms)	(Hz)
this study						
/(ʃ)i/	2070,	0,	2070	2530,	60,	2580
/(s)i/	1870,	80,	2050	2430,	140,	2600
/(ʃ)u/	1940,	300,	1000	2350,	60,	2150
/(s)u/	1620,	300,	1000	2340,	170,	2130
Whalen, 1981						
/(ʃ)i/	2100,	130,	2300	2750,	130,	2900
/(s)i/	1700,	130,	2300	2500,	130,	3000
/(ʃ)u/	1800,	300,	850	2100,	0,	2100
/(s)u/	1600,	300,	850	2750,	130,	2100

Figure 2: Spectrograms of vocalic portions used in this study



(sh)i (sh)u



(s)i (s)u

station permitted the adults to listen in groups of 4, and one of the experimenters to listen simultaneously with the child being tested. A Belltone Model 9D portable audiometer and a Grason-Stadler Model 27 portable tympanometer were used to screen pure-tone threshold and middle ear function.

Four hand-drawn pictures were prepared to correspond to each of the four possible responses. One picture was of a shoe; one was of a girl named 'Sue'; one was of a boy pointing and saying 'see'; and the last was of a girl referred to by the pronoun 'she'. Several board games were also prepared. These games simply consisted of brightly colored squares and circles (9 on each board) with the numbers 1 through 9 written on the spaces. A small plastic animal served as a marker, moving to the next space each time the female voice was heard announcing the number of the next group of 10 stimuli. Toy stickers, selected before the start of each game, were given as prizes at the end of the game.

Procedure

Each subject received the hearing screening before testing. Next, the children were shown the 2 pictures to be used in the first perceptual test (either 'see' and 'she', or 'Sue' and 'shoe'). Practice was provided using live voice to familiarize the child with the labeling procedure. This procedure consisted of having the child point to the picture illustrating the stimulus which was perceived and say the label associated with that picture. If there was a discrepancy between what the child seemed to be saying and the picture to which he/she was pointing, clarification was requested by the experimenter. After the child correctly responded to 10 practice items with live voice, the 10 taped practice items were presented.

If the child correctly responded to 9 out of 10 of these correctly, the board game was introduced, and the actual testing started. Stimuli were presented over earphones at approximately 75 dB SPL. Two experimenters were needed during the testing of the 3- and 4-year-olds: one to work with the child, listening and watching for discrepancies between his/her pointing and verbal responses as well as just maintaining attention with constant eye contact and occasional verbal praise, and one to stop the tape between stimulus presentations and record the child's responses. The experimenter working with the child was not able to hear the stimulus presentations at all. The experimenter recording children's responses wore the earphones around her neck so that she could hear when a stimulus had been presented, but could not hear it well enough to form her own impression of whether it was /s/ or /ʃ/. Thus, there was no danger of the response recorded being influenced by her own perception. Only one experimenter was needed for testing 5- and 7-year-olds.

Testing was divided into 4 sessions for the children; 2 sessions of 20-30 minutes duration, and 2 of 10-15 minutes duration. Activities during the sessions were arranged as follows:

Session 1: Hearing screening and first half of first perceptual test (1 test tape).

Session 2: Second half of first perceptual test, and first half of second perceptual test (2 test tapes).

Session 3: Second half of second perceptual test (1 test tape).

Session 4: Collection of production samples to be used in Experiment 2. (Details of this procedure will be provided in the next section.)

All testing was done in one day with breaks between sessions. Testing took place either in the home or in a daycare facility. The order of presentation of the perceptual tapes was counterbalanced across subjects within each age group. Half heard the /i/ tapes first, and half heard the /u/ tapes first. Adults simply responded by writing whether they heard 's' or 'sh'. In addition, adults listened in groups of 4, and all perceptual tapes were presented during one session, with the /i/ and /u/ tapes alternated.

RESULTS

Probit analyses were done on individual and group response proportions for each context. (The general term 'context' will be used here to refer to the 4 vocalic portions used: /(ʃ)i/, /(s)i/, /(ʃ)u/, and /(s)u/.) Probit analysis fits a cumulative normal curve to proportion scores as a function of stimulus level by the method of least squares (Finney, 1971), estimating the mean (phoneme boundary) and standard deviation for each distribution. The slope is obtained by taking the inverse of the standard deviation, and therefore, serves as an index of consistency[1]. Figures 3 through 7 display the identification functions obtained for each age group from probit analyses of mean proportions. Table 3 provides mean phoneme boundaries for the age groups. (Phoneme boundaries and slopes for individual subjects are provided in Appendices A and B, respectively. Note that the first digit of each subject number indicates age, and the second digit represents order of testing within age groups.) Lower phoneme boundaries indicate that more 's' responses were given to that context. Results should be compared (1) between different vocalic transitions for the same vowel (/(ʃ)i/ vs. /(s)i/ and /(ʃ)u/ vs. /(s)u/, and

Figure 3: Mean identification functions for 3-year-olds

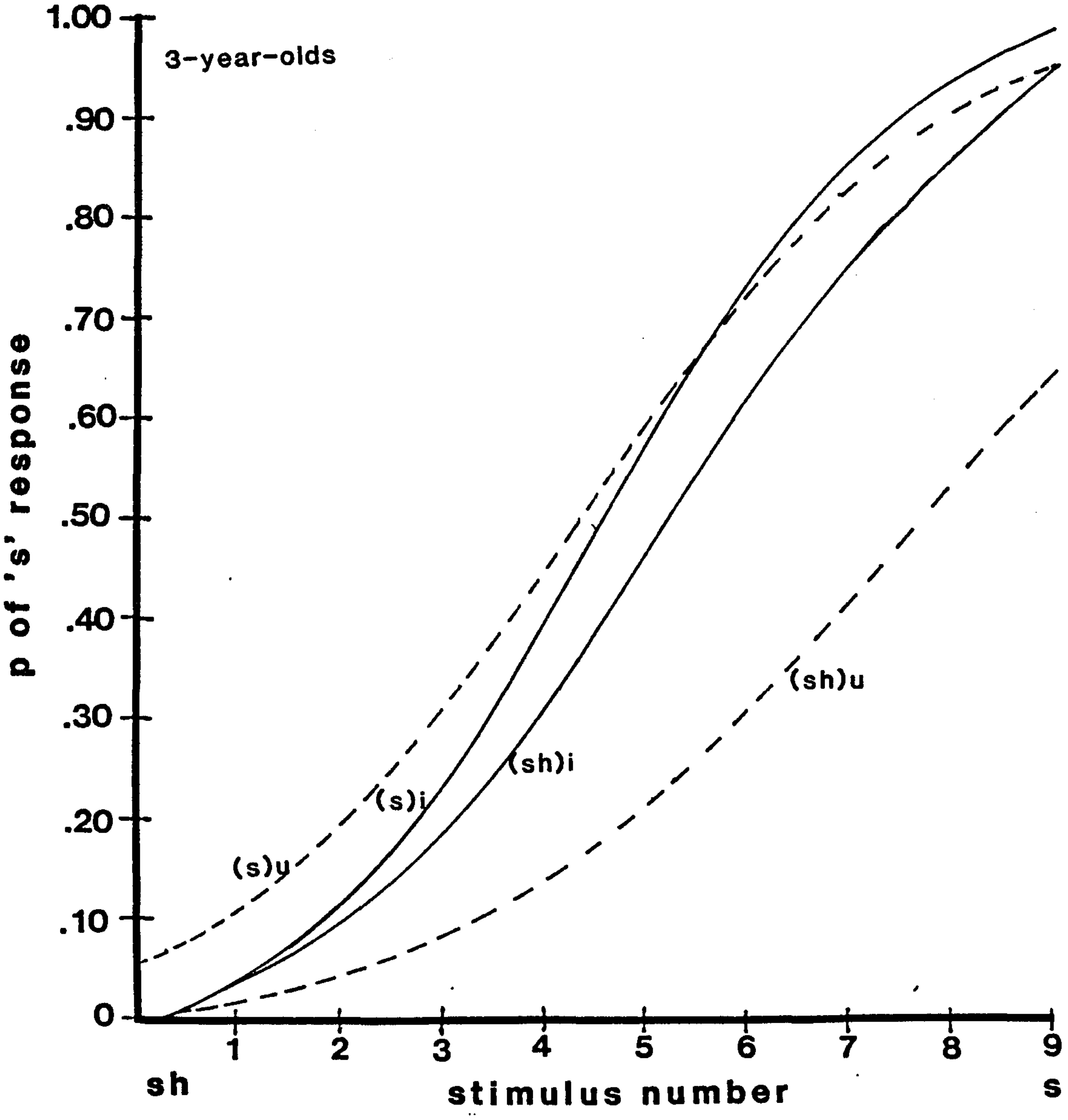


Figure 4: Mean identification functions for 4-year-olds

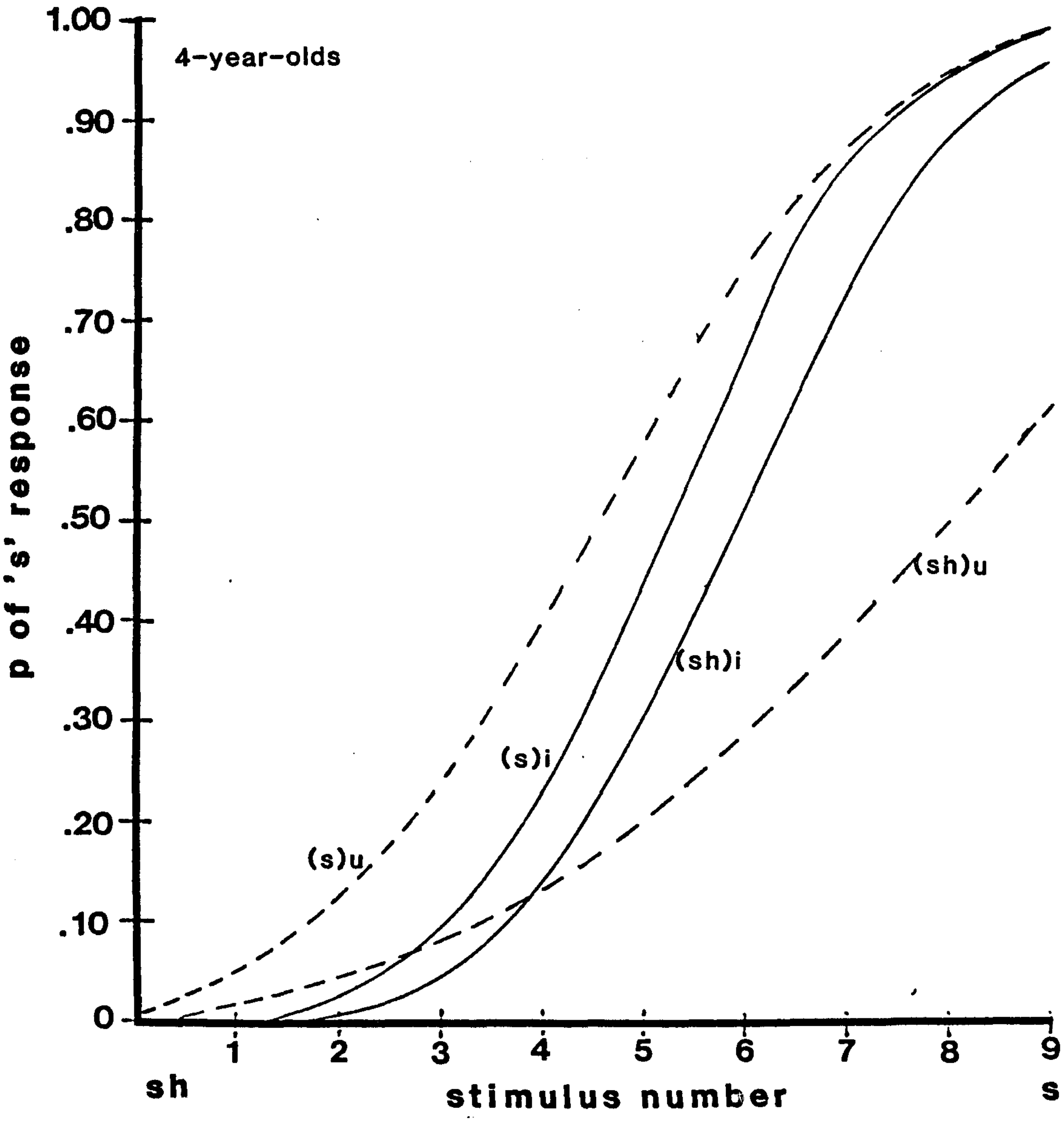


Figure 5: Mean identification functions for 5-year-olds

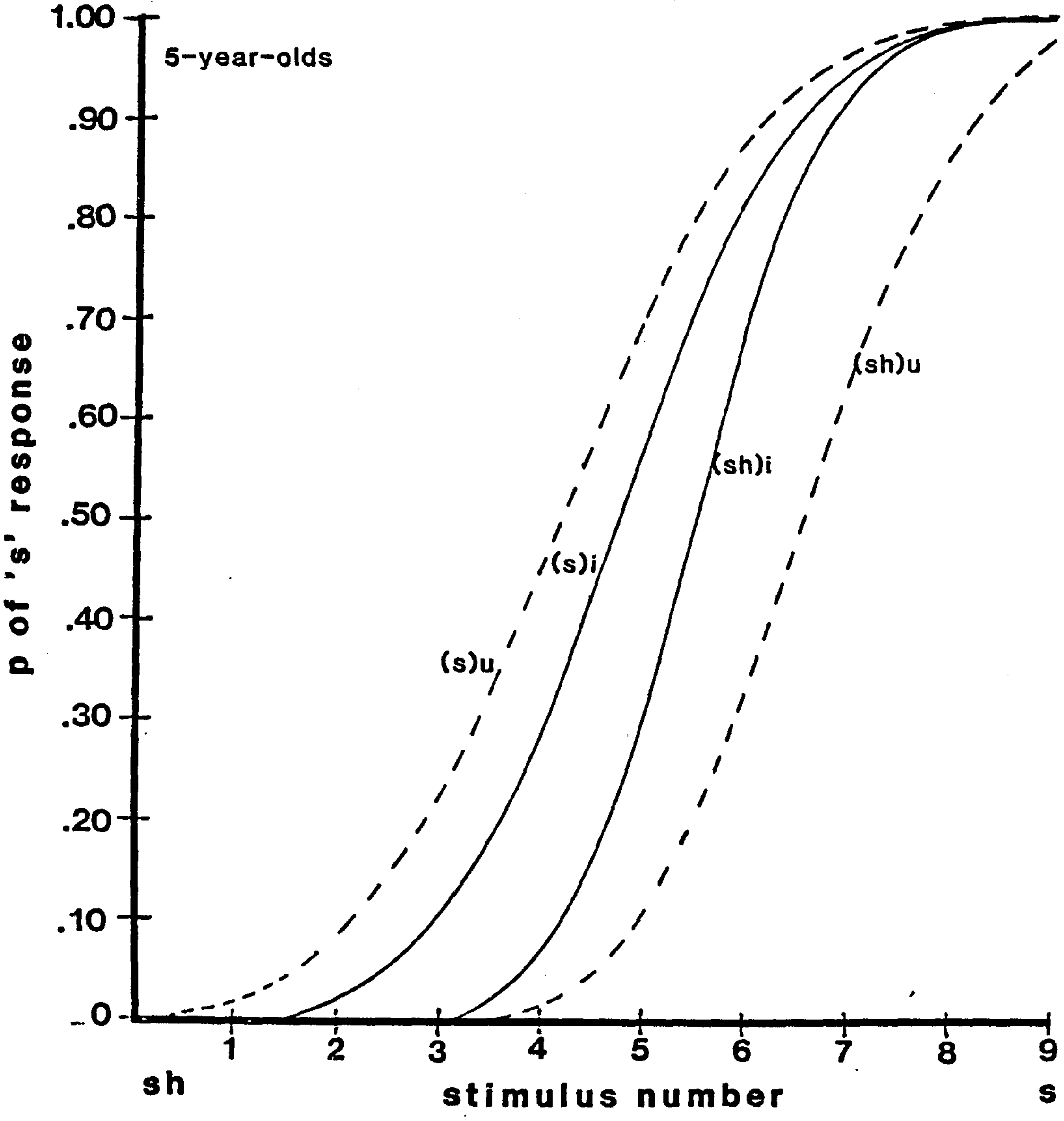


Figure 6: Mean identification functions for 7-year-olds

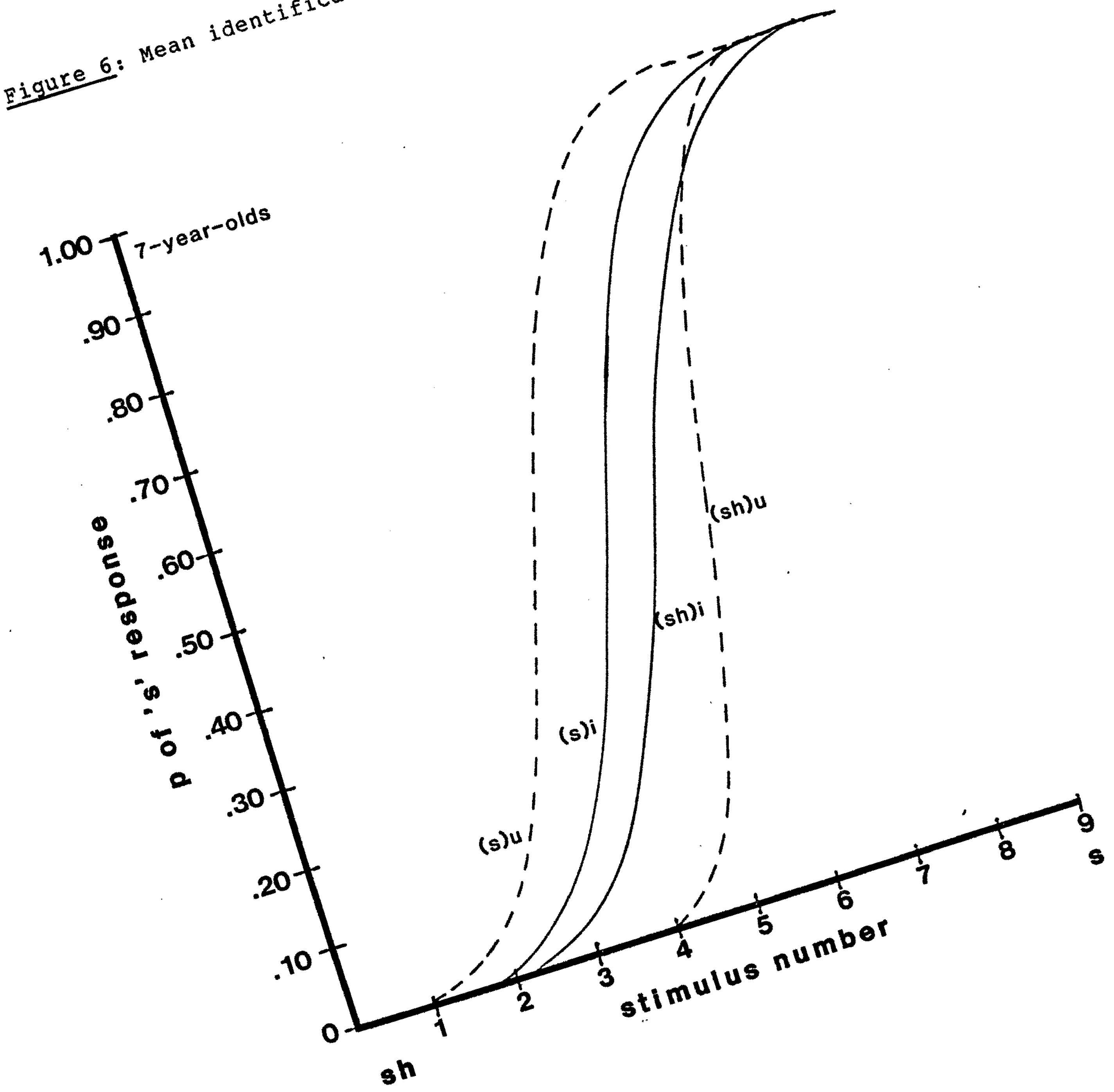


Figure 7: Mean identification functions for adults

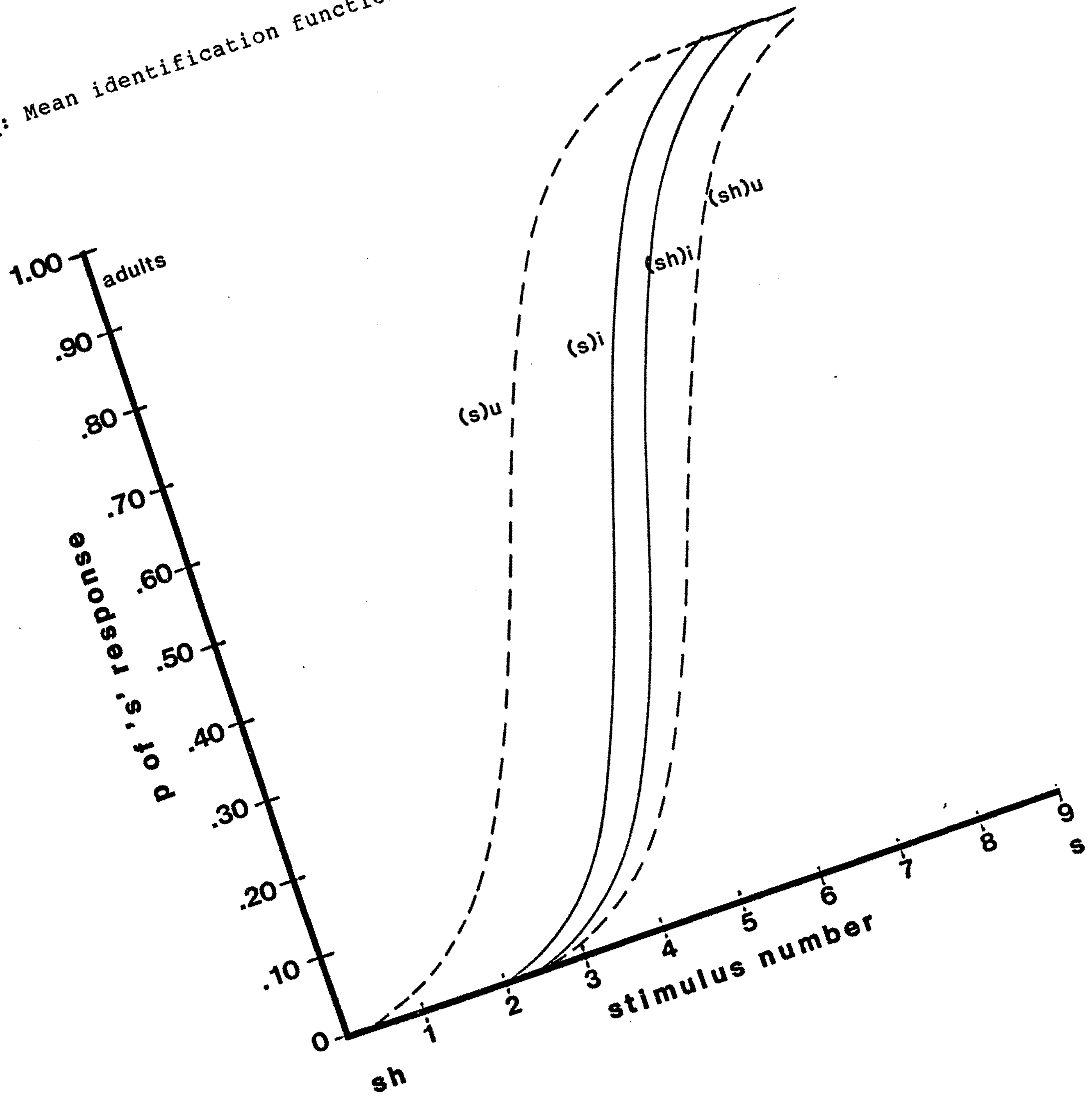


Table 3: Mean phoneme boundaries for each age group

	/(ʃ)i/	/(s)i/	/(ʃ)u/	/(s)u/
Adults	5.45	5.02	6.16	3.76
7-year-olds	5.21	4.63	5.81	3.74
5-year-olds	5.53	4.71	6.52	4.10
4-year-olds	5.95	5.22	7.89	4.51
3-year-olds	5.20	4.60	7.87	3.91

(2) between different vowel contexts for the same vocalic transition (/(ʃ)i/ vs. /(ʃ)u/ and /(s)i/ vs. /(s)u/). From the adult studies reviewed above, we would expect lower phoneme boundaries for tokens with /s/ transitions in the first case, and lower boundaries for /u/ contexts in the second.

Inspection of the mean phoneme boundaries given in Table 3 reveals that the predicted transition effect is apparent within both vowel contexts (lower phoneme boundaries for tokens with /s/ transitions), but the size of this effect appears greater for /u/ than for /i/. The expected vowel effect occurred only for tokens with /s/ transitions (a lower phoneme boundary for the /(s)u/ context than for the /(s)i/ context). For tokens with /ʃ/ transitions, the obtained vowel effect was opposite to what would be expected: a lower phoneme boundary was obtained in the /(ʃ)i/ context than in the /(ʃ)u/ context. In fact, all age groups demonstrated the highest phoneme boundary in the /(ʃ)u/ context, indicating more 'sh' responses in this context than any other.

A 2-way ANOVA (Age x Context) performed on individual phoneme boundaries revealed a significant main effect of context only (F(3,35)=76.2, p<.01). The main effect of age was not significant, however, the 2-way interaction of Age x Context was significant (F(12,105)=2.13, p=.02). Thus there was a difference in where subjects placed phoneme boundaries as a function of context, and the size of these differences varied across age groups.

While this analysis indicates that there was an age effect on where subjects placed phoneme boundaries for one or more context, it does not indicate for which specific context(s) this age difference occurred. To address this question, 1-way ANOVAs for age were performed on individual phoneme boundaries for each context. Only the analysis for the /(ʃ)u/ context revealed a significant effect of age (F(4,35)=3.61, p=.01). It seems that subjects of all ages placed their boundaries in approximately the same place for all contexts, except /(ʃ)u/ in which case younger subjects placed the phoneme boundary higher on the continuum indicating fewer 's' responses.

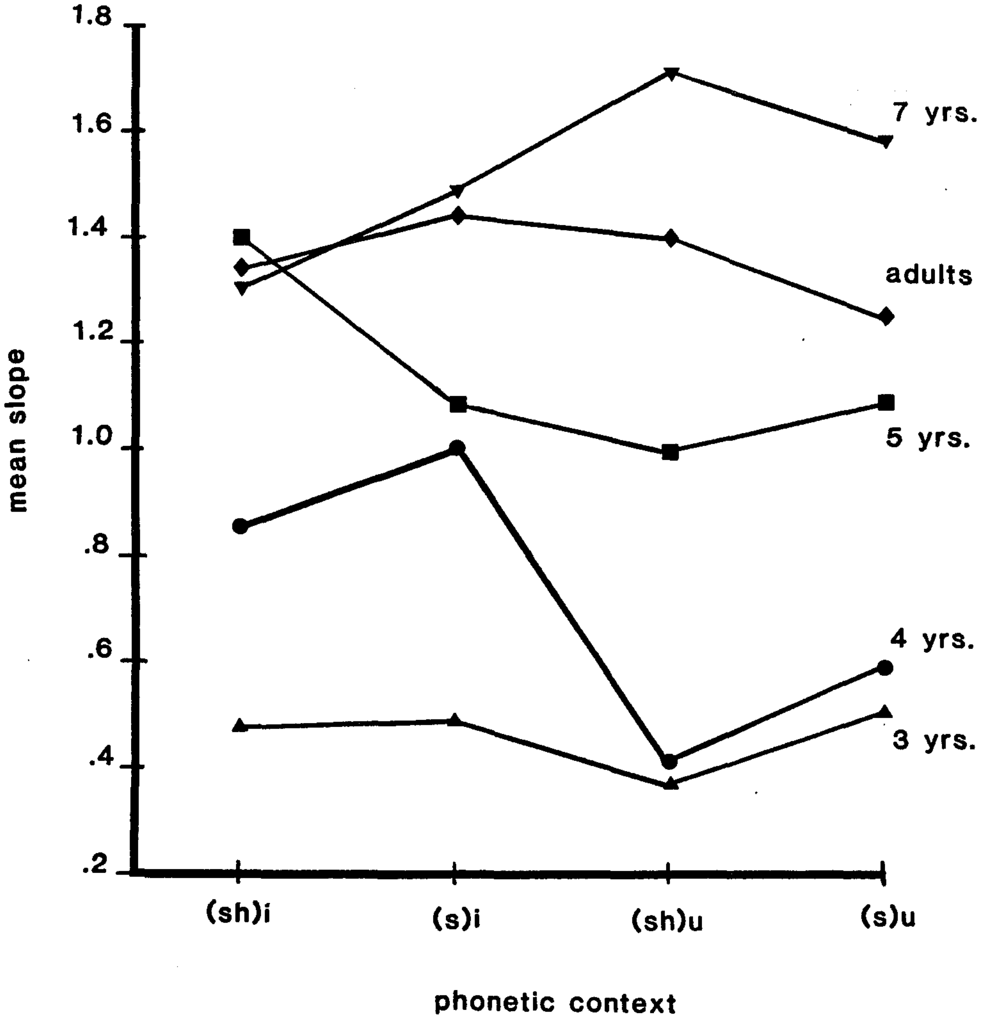
Inspection of Table 4 and Figure 8, illustrating slopes, reveals that slope values were generally greater in value for older subjects. Adults and 7-year-olds demonstrated mean slopes substantially above 1.00; slopes for 5-year-olds were closer to 1.00; and, with one exception, slopes for 4- and 3-year-olds were below 1.00.

Table 4: Mean slope values[a] for each age group

	/(ʃ)i/	/(s)i/	/(ʃ)u/	/(s)u/
Adults	1.34	1.44	1.41	1.26
7-year-olds	1.30	1.48	1.73	1.58
5-year-olds	1.40	1.09	1.01	1.08
4-year-olds	0.86	1.00	0.42	0.60
3-year-olds	0.47	0.50	0.37	0.52

[a] Slope values represent the change in probability of an 's' response per unit change in the fricative noise.

Figure 8: Mean slope values as a function of phonetic context for the 5 subject groups



Looking across contexts within age groups it can be seen that slope values were fairly consistent. Only 4-year-olds seem to display different slope values as a function of context, with the values for /i/ higher than the values for /u/. However, this difference is not statistically significant. The fact that slopes are higher for 7-year-olds than adults in the /u/ contexts may be attributable to Subject A4, who demonstrated individual slope values substantially lower in value than the other adults (see Appendix B).

A 2-way ANOVA (Age x Context) on individual slopes revealed a significant main effect of age ($F(4,35)=14.73$, $p<.01$). In general, older subjects exhibited steeper slopes --- that is, they were more consistent in their responses.

Difference Scores

Although subjects of different ages placed their phoneme boundaries in approximately the same location for 3 of the 4 contexts, it is the *relative* location of these boundaries for each subject which will provide information about the size of these effects. Therefore, the next step was to compute difference scores to serve as indices of the magnitude of vowel and transition effects. The difference in phoneme boundary for each of the 4 relevant comparisons was computed for each subject. Specifically, these difference scores were:

<u>Transition effects</u>

(ʃ)i-(s)i

(ʃ)u-(s)u

<u>Vowel effects</u>

(ʃ)i-(ʃ)u

(s)i-(s)u

Mean difference scores for each age group are given in Table 5 and illustrated in Figures 9 and 10. Individual difference scores are listed in Appendix D. Minus signs indicate that the difference was in the opposite direction from what would be predicted on the basis of previous studies.

It can be seen in Table 5 and Figures 9 and 10 that, overall, the magnitude of the transition effects is greater for the younger subjects. One exception to this trend is that the mean differences score for (ʃ)u-(s)u is greater for adults than for 7-year-olds. This result, again, is probably attributable Subject A4 who displayed a pattern of

Table 5: Mean phoneme boundary difference scores for each age group

	Transition Effects		Vowel Effects	
	(ʃ)i-(s)i	(ʃ)u-(s)u	(ʃ)i-(ʃ)u	(s)i-(s)u
Adults	0.43	2.40	-0.71[a]	1.26
7-year-olds	0.58	2.07	-0.60	0.89
5-year olds	0.77	2.42	-1.05	0.61
4-year-olds	0.72	3.38	-1.95	0.71
3-year-olds	0.61	3.95	-2.66	0.69

[a] Minus signs indicate that the difference is in the opposite direction from what would be expected based on earlier studies (Mann & Repp, 1980; Whalen, 1980)

Figure 9: Mean phoneme boundary difference scores by age group

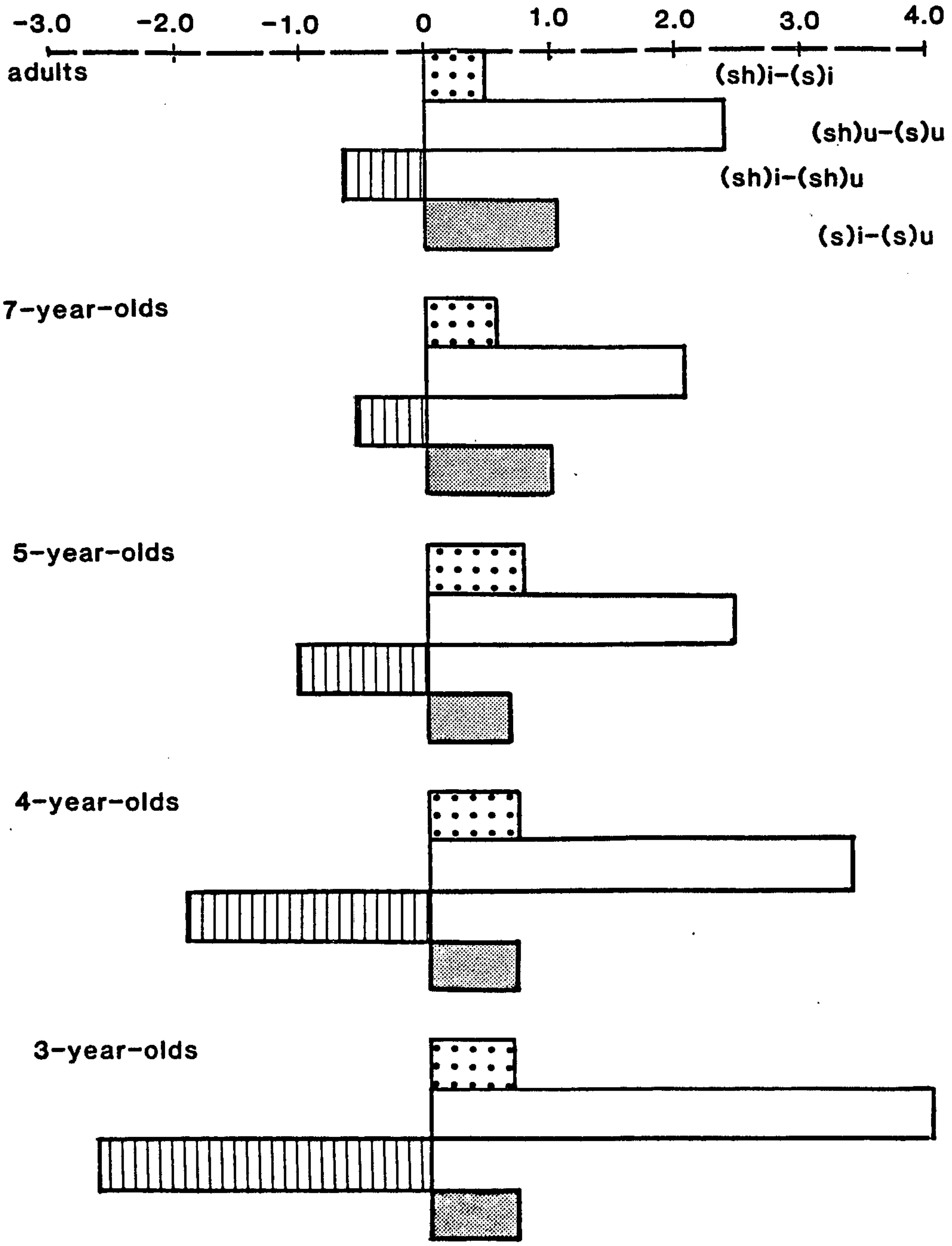
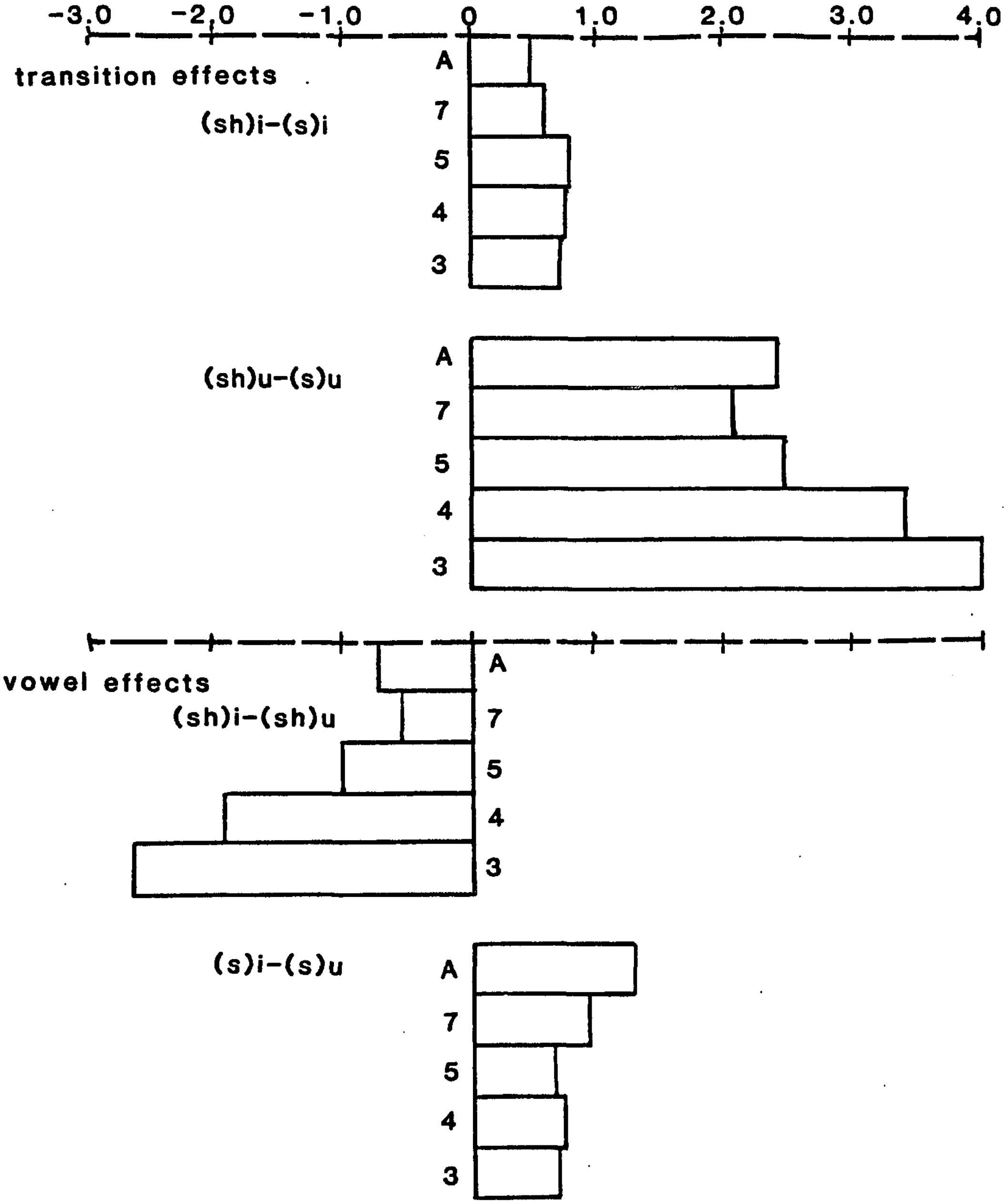


Figure 10: Mean phoneme boundary difference scores by effect type



results similar to that of young children: A large number of 'sh' responses was given in the /(ʃ)u/ context, resulting in a particularly large transition effect for /u/ tokens.

Considering vowel effects, Table 5 and Figures 9 and 10 show a tendency for older subjects to give larger difference scores for (s)i-(s)u than younger subjects. Difference scores for (ʃ)i-(ʃ)u indicate a vowel effect in the opposite direction of what would be expected for all age groups, and the size of this difference increases with decreasing age.

In this comparison ((ʃ)i-(ʃ)u)), 2 factors seem to have opposing effects on responses: the strong transition information provided by the /(ʃ)u/ context biases responses in favor of 'sh', while the vowel information provided by /u/ should, according to previous accounts, bias responses in favor of 's'. This difference is unlike that for (s)i-(s)u where both the vowel and the transition in the /(s)u/ context bias responses in favor of 's'. The size, then, of the (ʃ)i-(ʃ)u difference score really reflects the extent to which listeners were influenced by either the transition information or the vowel information. The closer this value was to zero, the stronger the vowel effect was. Thus it seems that older subjects were more influenced by vowel quality than younger subjects.

The questions that these difference scores could help answer are (1) whether there is a difference in magnitude between <u>types</u> of effects (i.e., vowel vs. transition effects), and (2) whether magnitude of effects varied as a function of age. A 2-way ANOVA (Age x Effect Type) performed on individual difference scores revealed no significant main effect of age, but a significant main effect of type ($F(1,35)=86.28$, $p<.01$). The magnitude of the transition effect was greater than the magnitude of the vowel effect. Furthermore, there was a significant interaction of Age x Effect Type ($F(4,35)=3.75$, $p=.01$). Younger subjects demonstrated relatively larger transition effects than vowel effects, as compared to older subjects.

In summary, then, children were less consistent in their responses than adults; children demonstrated smaller shifts in phoneme boundaries as a function of vowel quality than adults; however, children demonstrated larger shifts in phoneme boundaries as a function of formant transition than adults.

DISCUSSION

In considering whether children demonstrated coarticulatory effects to a greater or lesser extent than adults, three factors must be discussed: the vowel effect, the transition effect, and the extent to which fricative identification was based on the fricative spectrum itself. While each of these effects can be discussed as a separate phenomenon (and will be initially), it must be stressed that they are related because all three factors normally arise from articulation of the same syllable, so that their effects on perception would be expected to interact.

The expected vowel effect for these stimuli was that more 's' responses would be given to /u/ tokens than to /i/ tokens due to the lowering of fricative poles associated with rounded vowels. Children in this study demonstrated a weaker effect of this sort than adults. Visual inspection of the mean identification functions for each age group clearly shows a greater separation between the (s)i and (s)u functions for adults than for children, and for 7-year-olds than for younger age groups.

The transition effect for these stimuli predicted by earlier studies was that more 's' responses would be given to tokens with /s/ vocalic transitions. This effect was obtained for all age groups for both vowels, as illustrated in Figures 9 and 10, but the size of this effect was larger for /u/ than for /i/.

To make sense of these differences in transition effects between vowel contexts, it is necessary to look at the formant characteristics listed in Table 2, and illustrated in the spectrograms given in Figure 2. Not only is the difference in F2 onset values greater between /u/ tokens than /i/ tokens, but also the transitions are of longer duration in /u/ than in /i/. Thus, it seems that the /u/ tokens used in this study provided relatively more salient transitional information concerning fricative identity than the /i/ tokens.

Differences in the amount of transitional information provided by the different vowel contexts probably account for the differences in results found for this study compared to previous studies. The subjects in this experiment showed a pattern of results for tokens with /ʃ/ transitions different from that of subjects in earlier experiments (Mann & Repp, 1980; Whalen, 1981): These subjects gave more 's' responses to the unrounded vowel than to the rounded vowel for tokens with /ʃ/ transitions. Not only does this finding differ from results of earlier studies, it is contrary to the suggestion made by Kunisaki and Fujisaki (1977) that more 's' responses would be expected for rounded vowels than unrounded vowels, regardless of transition, because fricative poles are lowered preceding rounded vowels.

Second formant transitions are more similar for /a/ and /u/ than for /i/ and /u/ (Soli, 1981); therefore, the pattern of results obtained by Mann and Repp (1980) for tokens with /ʃ/ transitions is most likely a reflection of the vowel effect, to a large extent. That is, /a/ and /u/ are similar in the amount of transitional information provided, therefore, differences between contexts with vocalic transitions appropriate for the same fricative, but different vowels, would be largely due to vowel quality.

The /i/ portions used by Whalen (1981) exhibited greater transitional differences than the /i/ portions used in this study: Differences in F2 onset values were twice as large between /(ʃ)i/ and /(s)i/, and the extent of /i/ transitions were greater. Additionally, the acoustic differences between /(ʃ)u/ and /(s)u/ portions used by Whalen do not appear as great as the differences between the /u/ portions of this study. In particular, the /(ʃ)u/ portion used by Whalen does not appear to provide as much transitional information as the /(ʃ)u/ portion used here because of the fact that F3 shows no change in Whalen's /(ʃ)u/. In short, the amount of salient transitional information provided by /i/ and /u/ tokens seems more similar for Whalen's stimuli than for these stimuli. Therefore, differences in placement of /(ʃ)i/ and /(ʃ)u/ boundaries would be more a function of vowel context than of differences in amounts of transitional information provided by each context.

It might be hypothesized that the perceptual weight given to transitions is directly proportional to the amount of formant movement. The transitions of the /(ʃ)i/ portion used here displayed no formant movement and may therefore represent ~~the~~ neutral condition; /(s)i/ had weak transitions favoring 's'; /(s)u/ had somewhat more perceptually salient transitional information favoring 's'; but /(ʃ)u/ provided very strong transitional information favoring 'sh'. Superimposed on this pattern seems to be an independent vowel effect that tends to move the two /u/ functions in the direction favoring 's' responses --- at least for adults who demonstrate sensitivity to vowel quality. Clearly, the relation between the size of transitional differences found in acoustic measurements and the size of boundary shifts observed in perception warrants more systematic investigation. However, regardless of what the precise nature of this relation turns out to be, the finding that subjects in the present experiment gave more 's' responses to the /(ʃ)i/ context than to the /(ʃ)u/ context argues against explanations of the vowel effect based primarily on auditory contrast. If the predicted vowel effect were a consequence of the fricative noise sounding higher when the vowel spectrum is lower, as in /u/ compared to /i/, this effect should be found regardless of transition.

There was also an effect of age on the size of the transition effect: younger subjects demonstrated greater transition effects than older subjects, particularly for /u/. This age effect seems to be largely (though not entirely) a function of younger subjects giving substantially more 'sh' responses to the /(ʃ)u/ context than older subjects, as indicated by the statistically significant age effect for phoneme boundaries in the /(ʃ)u/ context only.

These /(ʃ)u/ tokens presumably had two opposing factors influencing the locations of phoneme boundaries: the percepually salient transitional information biased responses toward 'sh', and the rounded vowel biased responses toward 's'. The fact that younger subjects gave substantially more 'sh' responses to /(ʃ)u/ tokens than older subjects seems to be another indication that younger subjects were more greatly influenced by transitional information than older subjects, and less influenced by vowel quality. While /(s)u/ tokens provided perceptually salient transitional information concerning fricative identity, vowel quality was biasing responses in the same direction (toward 's' responses). Therefore, it is impossible to determine relative contributions of vowel and transition for responses in this context.

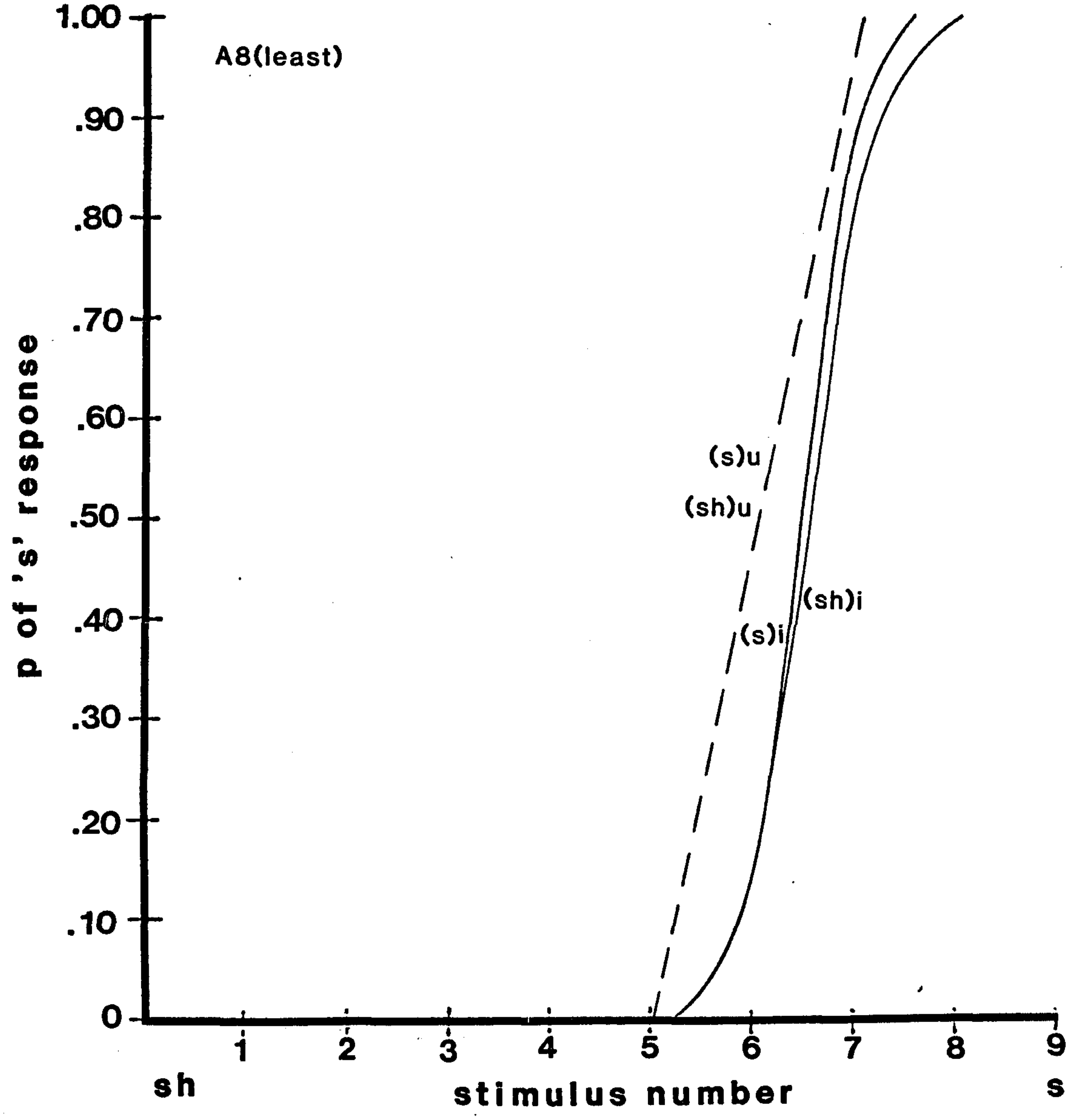
Based on these results, it appears that vowel effects and transition effects are additive. The negative boundary shift obtained for the /(ʃ)i-(ʃ)u/ difference evidently reflects the overwhelming cancellation of the vowel effect by the /(ʃ)u-(s)u/ transition effect, i.e., vowel and transition effects are orthogonal. The fact that the 3- and 4-year-olds demonstrate such a large negative vowel effect for /ʃ/ tokens is actually a result of their demonstrating a very large transition effect for /u/.

Another purpose of the present experiment was to investigate the extent to which fricative identification was associated with the fricative spectrum itself. To make this determination, it is necessary to look at the slopes of the identification functions. The finding that 3-, 4-, and 5-year-olds exhibited shallower slopes than 7-year-olds and adults shows that they were less consistent in responding. This lack of consistency may be associated with the children being less attentive to the fricative noise. One subject who did appear to attend selectively to the fricative noise, and to make decisions about category based primarily on its spectrum was Subject A8, whose identification functions are shown in Figure 11. These functions show very consistent responses and no effect of vocalic transition; however, they do exhibit a vowel effect.

This last result is important because it suggests that attention to the fricative noise is associated with a vowel effect. Vowel context effects are found because the fricative spectrum differs as a function of the quality of the following vowel. If a listener does not attend to the fricative noise to some extent, then it follows that he/she would not be able to determine whether this noise differs as a function of the following vowel.

This relation between attending to the fricative noise and showing vowel effects might explain why young children did not demonstrate vowel effects to the same extent as adults in this study. Morrongiello et al.

Figure 11: Identification functions for Subject A8 (least)[a]



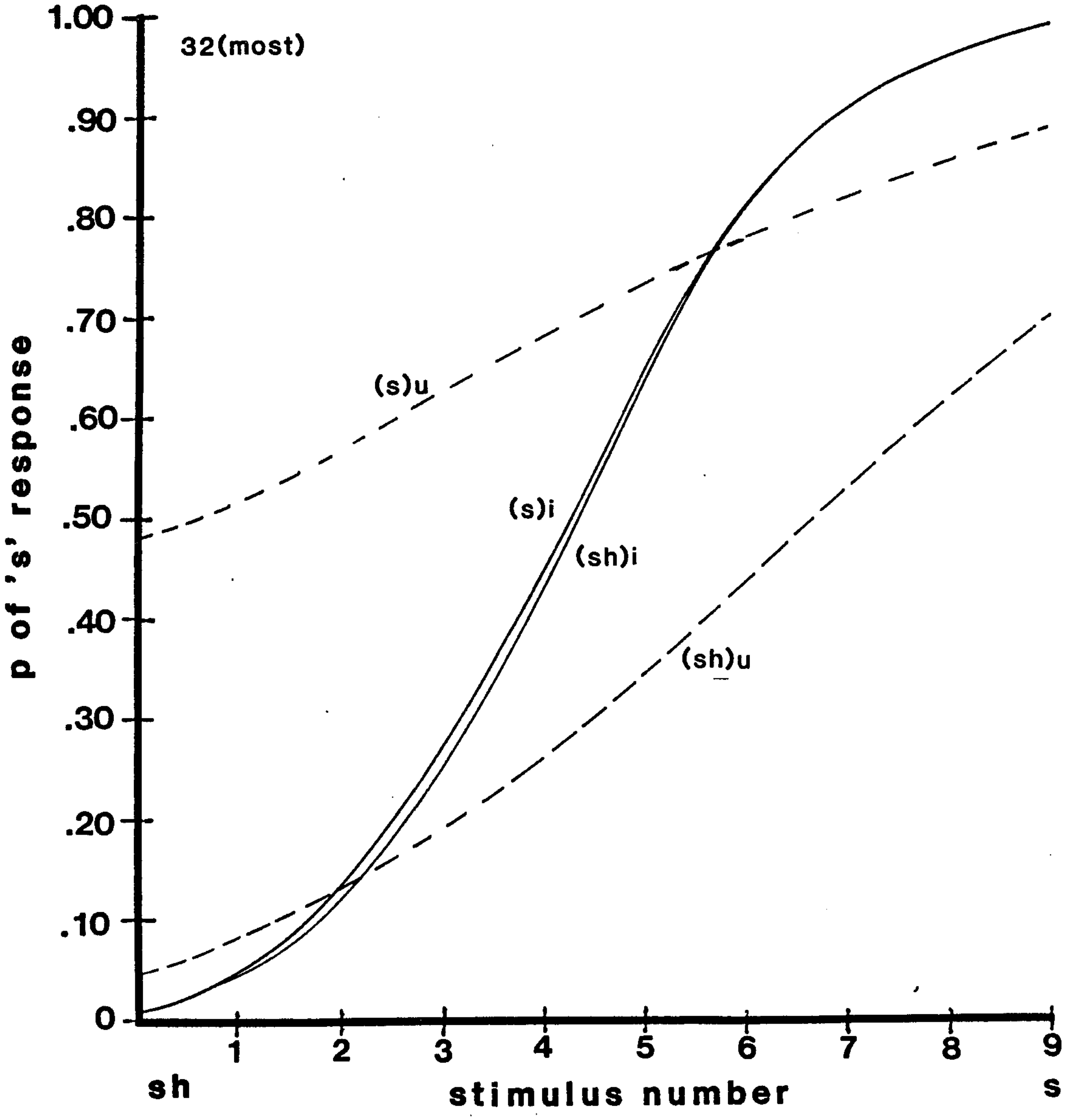
[a] The 'least' in parentheses indicates that this subject demonstrated the least effects of context on fricative identification for this age group.

(1984) reported that the children in their study appeared to weight the transitional information relatively more heavily than the silent interval, as compared to adults. That is, children seemed to pay relatively more attention to the transitional pattern than to the silent interval. The results obtained in the present experiment show a similar pattern: children apparently attended more to the transitional pattern than to the fricative noise, as compared to adults.

The shapes of the identification functions also indicate the increased reliance on transitional information shown by the children in this study. If categorical decisions were more heavily weighted by vocalic transitions than by fricative noise, it would be expected that responses would be biased in favor of the fricative associated with that transition, and that many of the stimuli would be ambiguous. Therefore, phoneme boundaries would be extremely low or high, and slope values would be low.

One example of a subject whose responses to /u/ tokens seem to have been controlled primarily by transitional information is Subject 32, whose identification functions are shown in Figure 12. Phoneme

Figure 12: Identification functions for Subject 32 (most)



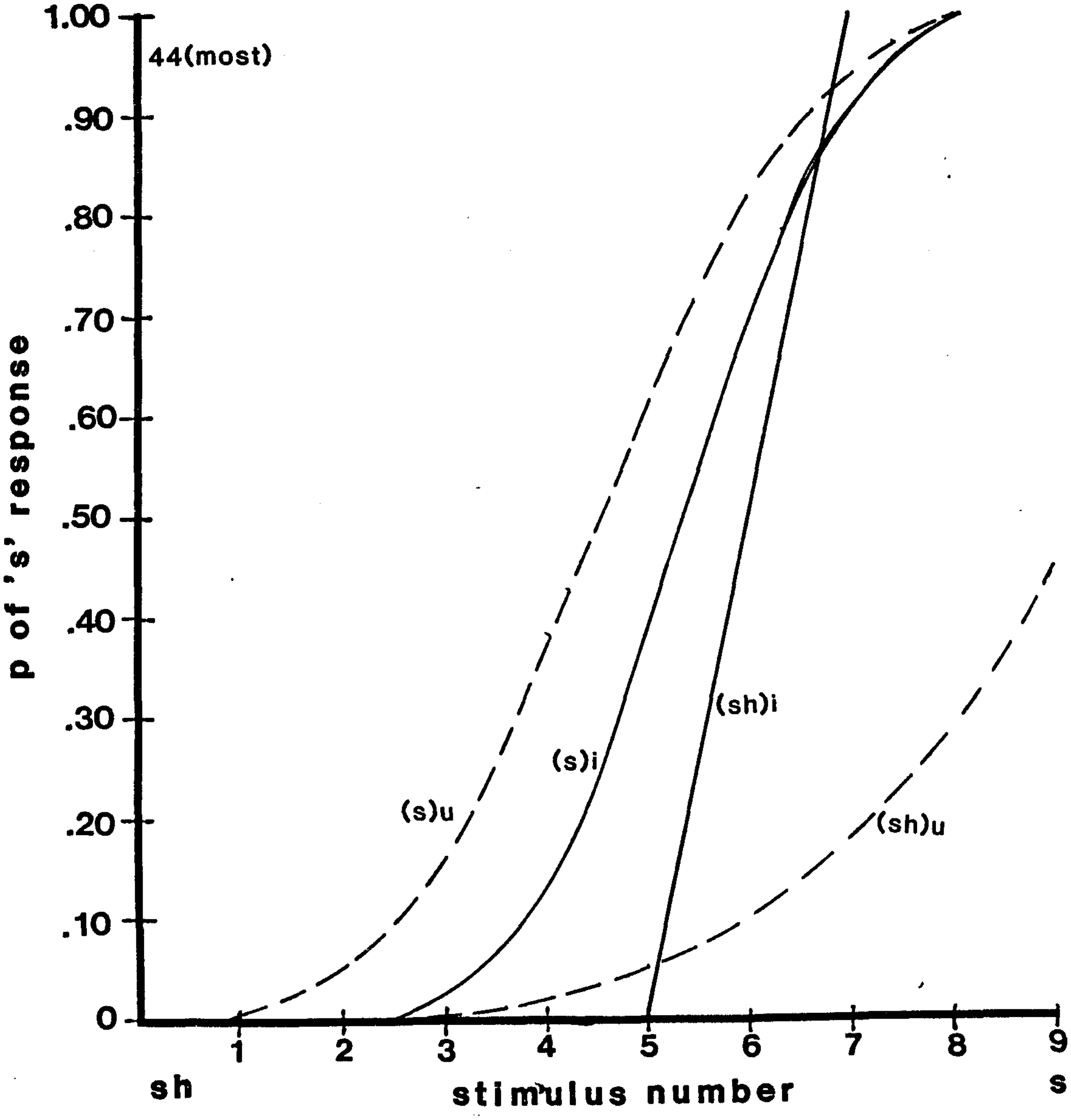
boundaries are extreme for /(ʃ)u/ and /(s)u/, and slope values are low for these contexts. Similar patterns (extreme phoneme boundaries and low slope values) can be found in results obtained for some of the other 3- and 4-year-olds (and for Subject A4), but only for /u/ contexts. No subject displayed a pattern of results indicating strong influence of transitions in either of the /i/ contexts, probably because /i/ transitions provided less salient information about fricative identity than did /u/ transitions. However, many of the 3-year-olds displayed very low slope values, simply indicating inconsistent responses for the /i/ contexts. Apparently, even when transitional information was greatly reduced, 3-year-olds were not able to increase their use of the cues provided by the fricative spectrum for fricative identification.

Four-year-olds seem to have been able to switch response patterns to some extent depending on how much information was provided by the transitions. For the /u/ contexts in which the transitions provided perceptually salient information concerning fricative identity, responses seem to have been based largely on vocalic transitions, as indicated by the low slope values and extreme phoneme boundaries. For the /i/ contexts in which transitions provided little information, they seem to have been able to turn their attention to the fricative noise and use its spectrum to make decisions about fricative identity to some extent, as indicated by higher slope values.

A good example of this response pattern is Subject 44, whose identification functions are shown in Figure 13. Responses for the /(ʃ)u/ context are strongly influenced by the vocalic transitions, as indicated by the shallow slope and extremely high phoneme boundary, but responses for the /(ʃ)i/ context are clearly based on the fricative spectrum, as indicated by the very steep slope.

Another developmental trend found in these data is that 7-year-olds performed similarly to adults in most respects. Their responses were as consistent as adults', indicating strong reliance on the fricative spectrum for fricative identification, and the magnitude of the transition effect for individual 7-year-olds was equivalent to that obtained for individual adults. The only major difference between adults and 7-year-olds was in the magnitude of the vowel effect: 7-year-olds demonstrated lower difference scores for these effects. In general, though, 7-year olds appear to have developed speech perception strategies similar to adults'.

Figure 13: Identification functions for Subject 44 (most)



Finally, all results reported here are in agreement with the integrating account of how acoustic cues are used in speech perception (Whalen, 1984). The 3- and 4-year-olds seem to have been more strongly influenced by vocalic transitions than were older children and adults. According to the disposing account, the children should have demonstrated an even stronger reliance on the fricative spectrum for responding than adults. The fact that they did not contradicts the disposing account, in which transitional information is viewed as "secondary", and static target information as "primary".

Contextual effects in children's speech perception were initially studied with the expectation that the effects of context on perception would be less than or equal to those of adults. Instead, these effects appear to be generally greater in children's speech perception. However, it also appears that young children are less capable of attending to the specific portions of the syllable primarily associated with individual segments; i.e., fricative and vowel. This finding may reflect a whole word or syllable approach to speech, such as that described by Menyuk and Menn (1979) for 1- and 2-year-old children. However, the idea that children as old as 3 and

4 perceive speech as whole syllables or words is contradicted by the size of their vocabularies. It would be very difficult to store enough whole syllable or word templates to permit a vocabulary as large as those demonstrated by 3- and 4-year-olds, as researchers in automatic speech recognition are discovering (Levinson & Liberman, 1981). Alternatively, it could be that 3- and 4-year-olds are beginning to discover the phonetic segments packaged in the syllable, but are not yet as adept at recovering them as adults. They may "know" that segments are there, but are not yet able to retrieve them fully.

This suggestion that young children perceive speech as patterns larger than the phoneme is also compatible with the notion that, for adults, different segments preserve their qualitative separateness (Fowler, Rubin, Remez, & Turvey, 1980). If we accept that vowel and consonant articulation result in qualitatively different acoustic events, then it would seem from these results that children are beginning to distinguish between the acoustic events of vowel and consonant articulation at 3 and 4, but are not quite able to do this successfully until approximately 7 years of age.

Finally, the results reported for this experiment seem consistent with the notion that children would attend to the obligatory consequences of coarticulation earlier than to the optional consequences. Children in this study appear to have weighted the essential coarticulatory effect (i.e., formant transition) relatively more heavily than the optional effect (i.e., lowering of fricative poles due to anticipatory liprounding). However, this distinction between essential and nonessential coarticulatory effects cannot adequately account for the pattern of results obtained by Morrongiello et al. (1984) because it seems that both formant transitions and a silent interval would necessarily result from the production of a fricative-stop-vowel syllable.

Experiment 2: Production

REVIEW OF THE LITERATURE

Hughes and Halle (1956) analyzed the spectra of fricative noises produced by one female and two male speakers in various vowel environments, and found the same general pattern for all: the most prominent energy peak was lower in /ʃ/ than /s/. In a further analysis of the fricatives produced by one of these male speakers, Heinz and Stevens (1961) reported that different samples of both /ʃ/ and /s/ displayed similar spectral shapes consisting of two poles and one zero. Although spectral values for the actual speech samples were not given, the poles and zeros of the transfer functions used to synthesize good matches to the natural word samples were provided. Mean settings in Hertz for /ʃ/ and /s/ are listed in Table 6. In addition, the values provided by Heinz and Stevens indicate that one of the poles in each fricative had to be set lower before rounded vowels than before unrounded vowels to synthesize good matches. The second pole of /ʃ/ was set at 4300 HZ for 'sure' as compared to 5400 Hz for 'sheep'. The first pole of /s/ was set at 3500 Hz for 'soothe' as compared to 6400 Hz in 'sect'. However, the

Table 6: Mean settings (in Hz) for transfer functions used by Heinz and Stevens (1961)

	first pole	second pole	zero
/ʃ/	2380	4983	3767
/s/	4850	8083	2750

pole which was set lower in frequency was always 5 to 10 dB below the other pole, so that the effects of this lowering on the overall spectrum were diminished. The most prominent pole in each, then, remained fairly constant in frequency across vowel contexts. This prominent, constant peak was always the lower pole for /ʃ/, and the higher pole for /s/. Moreover, it is not uncommon for there to be discrepancies between the values required for synthesis and those found in natural speech. In fact, Heinz and Stevens further demonstrated that although two poles and one zero are typically found in naturally produced fricatives, one pole and one zero are sufficient to produce perceptually adequate exemplars of /ʃ/ and /s/ in synthesis. Thus the extent to which vowel context affects fricative noise spectra cannot be determined from their study.

Strevens (1960) described amplitude sections taken from two locations in fricatives produced in isolation by 13 speakers. He found that the lower limit of noise for /s/ was consistently higher than 3500 Hz, and that the upper limit exceeded 8000 Hz. According to Strevens, the frequency of the major spectral peaks for /s/ displayed no obvious patterns. For /ʃ/, the lower limit of noise was found to vary between 1600 Hz and 2500 Hz with the upper limit at approximately 7000 Hz. Major spectral peaks consistently appeared toward the lower end of the noise spectra for /ʃ/. Because only fricatives produced in isolation were analyzed in this study, no information concerning coarticulatory effects was provided. Instead, this study is important because it is the only production study to analyze fricative noises for more than a few subjects, and therefore provides normative descriptions of the general patterns of /ʃ/ and /s/ spectra.

Kunisaki and Fujisaki (1977) compared the first pole and zero values of /ʃ/ and /s/ samples produced by a single male speaker in the medial position of VCV disyllables. Because various vowel contexts were used, some data were collected concerning coarticulatory effects in fricative production. In general, first pole and zero values tended to be lower when the following vowel was rounded than when it was unrounded. However, no measurements were made either on the second pole or on the overall spectral characteristics of the fricative, so that it is not possible to draw any comprehensive conclusions concerning coarticulatory effects.

The studies described above all support the acoustic theory of fricative production, as described by Fant (1960). This theory proposes that in fricative production, a noise source is created by air flowing through a narrow constriction made at some point along the vocal tract. According to this theory, only cavity resonances in front of this constriction contribute to the output spectrum because the constriction is sufficiently narrow to prevent coupling of the back cavity to the source. These studies also suggest that liprounding, by lengthening the front cavity, tends to lower fricative spectra. However, none of them considered the influence of precise tongue shaping and movement on output spectra.

Soli (1981) argued that, while Fant's acoustic theory of fricative production is adequate for describing steady-state portions of fricatives some distance from adjacent segments, other factors must be considered when the vocal tract volumes are changing, as in coarticulated production of a fricative and a neighboring vowel. First, the characteristics of a following vowel may influence the precise location and length of the constriction, if the articulatory components of the two segments are compatible --- that is, if vowel articulation can be initiated early enough in fricative articulation. Second, as the speaker approaches the next segment, the constriction necessarily begins to open, allowing the back-cavity resonances to influence the spectrum. Soli therefore computed mean LPC spectra of four fricatives, in three vowel contexts (/i/, /u/, and /a/) and in isolation, spoken by one male speaker (himself), to determine whether or not spectral peaks could be found corresponding to the formants of the following vowel. These spectra were computed in 10 ms time slices from a point 60 ms before vowel onset to the point of vowel onset.

Spectra obtained by Soli for /si/, /su/, /ʃi/, and /ʃu/ show that formant peaks corresponding to F2, F3 and F4 generally begin to emerge in the fricative spectrum 30 to 60 ms (or more) before vowel onset, and that these peaks are more prominent in /ʃ/ than /s/. No peaks are apparent in the /a/ context, presumably because the articulation of a low vowel is incompatible with /s/ and /ʃ/ articulations, which require high tongue positions. Therefore, articulation of /a/ cannot be initiated as early in fricative production as articulation of /i/ and /u/.

Formant trajectories were also obtained by Soli for a section starting at 60 ms before vowel onset and extending 100 ms into the vocalic portion. The fricative F2 values obtained by Soli, 60 ms before the onset of /i/ and /u/, are different; but the differences are slight and remain constant until approximately 20 ms before vowel onset. At this point, the differences in F2 values for the /i/ and /u/ contexts increase sharply. Soli suggests that this increased difference in F2 values 20 ms before vowel onset, as a function of vowel context, may indicate the increasing influence of the back cavity as the constriction begins to open in anticipation of the vowel. In fact, the influences of the back cavity (of which the volume is smaller for /ʃ/ than for /s/) can be inferred from the higher F2 values at onset for a vowel after /ʃ/ than for a vowel after /s/ (see, for example, Table 2).

Soli's study is important, first because it demonstrates that information about vowel identity is present in the fricative noise in the form of spectral peaks; second because it suggests that not only anticipatory liprounding, but also anticipatory tongue shaping and movement may affect the acoustic structure of fricatives.

Both Mann and Repp (1980) and Whalen (1981) measured formant frequency values at vocalic onset in /si/, /su/, /ʃi/, and /ʃu/. These values, indicate that F2 and F3 are higher at onset following /ʃ/ as compared to /s/, as Soli's argument predicts, and for /i/ as compared to /u/. In addition, Whalen's data show that the duration and extent of vocalic transitions are somewhat greater for /u/

than for /i/, presumably reflecting the more extensive tongue movement from palatal constriction into the high back vowel /u/ than into the high front vowel /i/.

The results described above demonstrate that information about vowel identity is present towards the end of the fricative noise in the form of spectral peaks in the region of F2, -- a finding compatible with the coproduction of segments posited by Fowler et al. (1980). However, the formant peaks in a fricative spectrum are difficult to resolve instrumentally, and we have no clear evidence that they are resolved perceptually in fricative identification. We may reasonably suppose, in fact, that the perceptually effective property of a fricative is less its formant structure than its overall spectral weight.

One goal of the present experiment was to obtain a measure of the overall fricative spectrum, and to determine if this measure varies as a function of vowel context. To confirm earlier studies, F2 values were also analyzed (1) in the fricative to see if there was localized information about the following vowel, and (2) in the vowel to see if there was localized information about the preceding fricative. Finally, production samples of adults and children were compared to determine whether or not children coarticulate to the same extent as adults.

METHOD

Subjects

Production samples were analyzed for all the adults who participated in the perception experiment to provide a baseline for comparison with the child data, and for the two children in each age group who demonstrated the greatest and the least contextual effects in perception in order to see if there was any relation between the size of these effects in perception and in production. These subjects were chosen largely on the basis of their difference scores for the transition effect (ʃ)u-(s)u and the vowel effect (s)i-(s)u. In most cases, the choices were clear: The subject who demonstrated the greatest or the least effect for one of these scores did so for the other as well. The only age group for which the choices were not obvious were the 4-year-olds who demonstrated greater intrasubject variability in patterns of results. For example, Subject 42 demonstrated the largest effect for (ʃ)u-(s)u, but the second least for (s)i-(s)u. A choice as to which 4-year-olds to include in the production analyses was eventually made by consideration of all four difference scores. The subjects chosen for acoustic analyses are listed in Table 7. Individual identification functions are included for the subjects in each age group demonstrating the greatest and the least perceptual sensitivities to coarticulatory effects. Most of these functions are provided in Figures 14 through 20. However, identification functions for three of these subjects (A4, 32, and 44) were already provided in the discussion section of the perception experiment (Figures 11 through 13).

The adults were rank-ordered according to the size of contextual effects demonstrated in the perception experiment for (ʃ)u-(s)u and (s)i-(s)u differences. So, for example, while Subject A7 showed a slightly greater effect for (ʃ)u-(s)u than Subject A3, Subject A3 received a higher ranking because (s)i-(s)u was more than twice as large for Subject A3 than for Subject A7. These ranks are shown on Table 7.

Table 7: Subjects selected for production experiment

Children

	Greatest Effect	Least Effect
7-year-olds	71	72
5-year-olds	52	56
4-year-olds	44	41
3-year-olds	32	35

Adults

	Rank	Subject
Greatest Effect	1	A4
	2	A1
	3	A6
	4	A3
	5	A7
	6	A5
	7	A2
Least Effect	8	A8

Figure 14: Identification functions for Subject 35 (least)

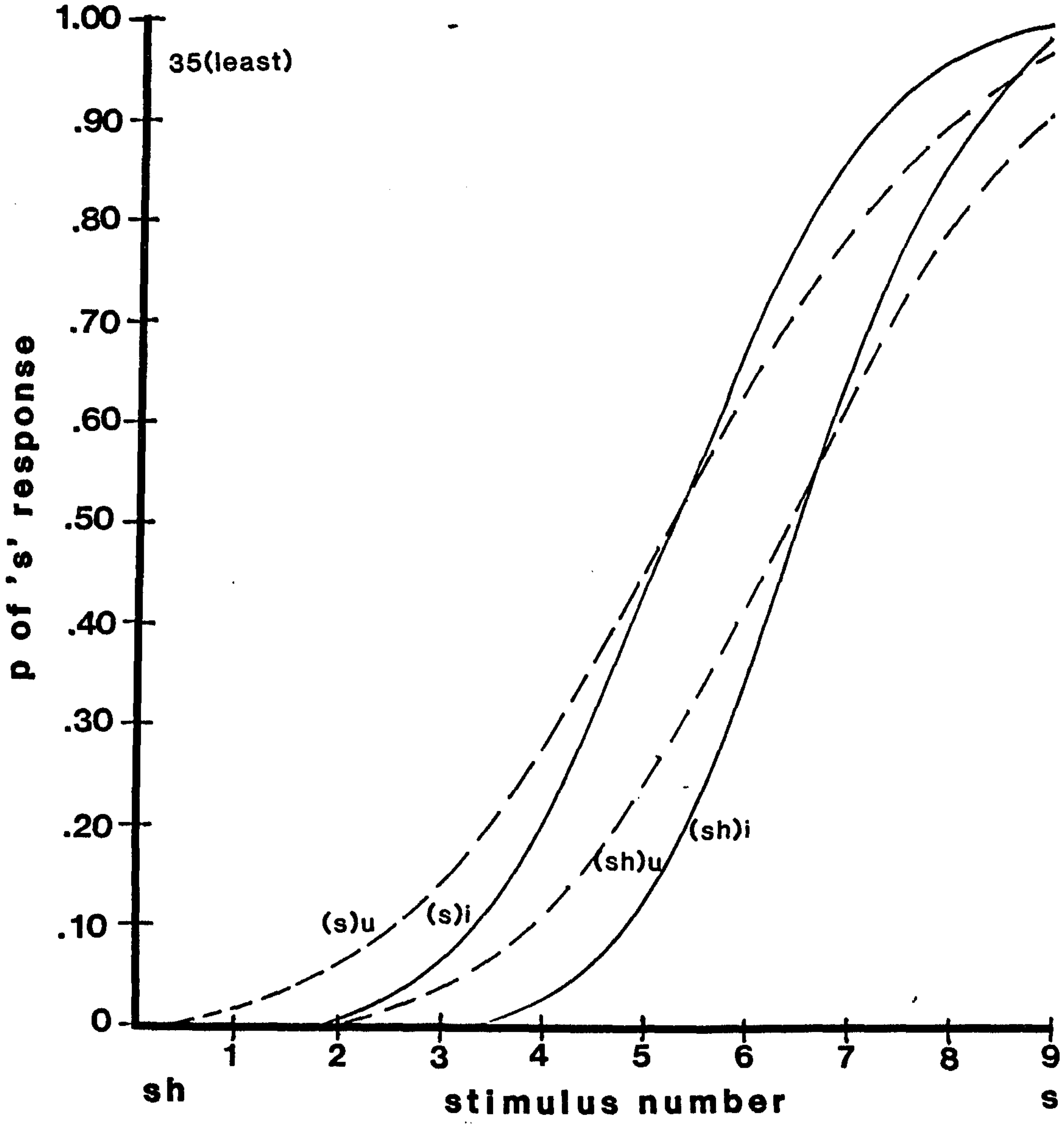


Figure 15: Identification functions for Subject 41 (least)

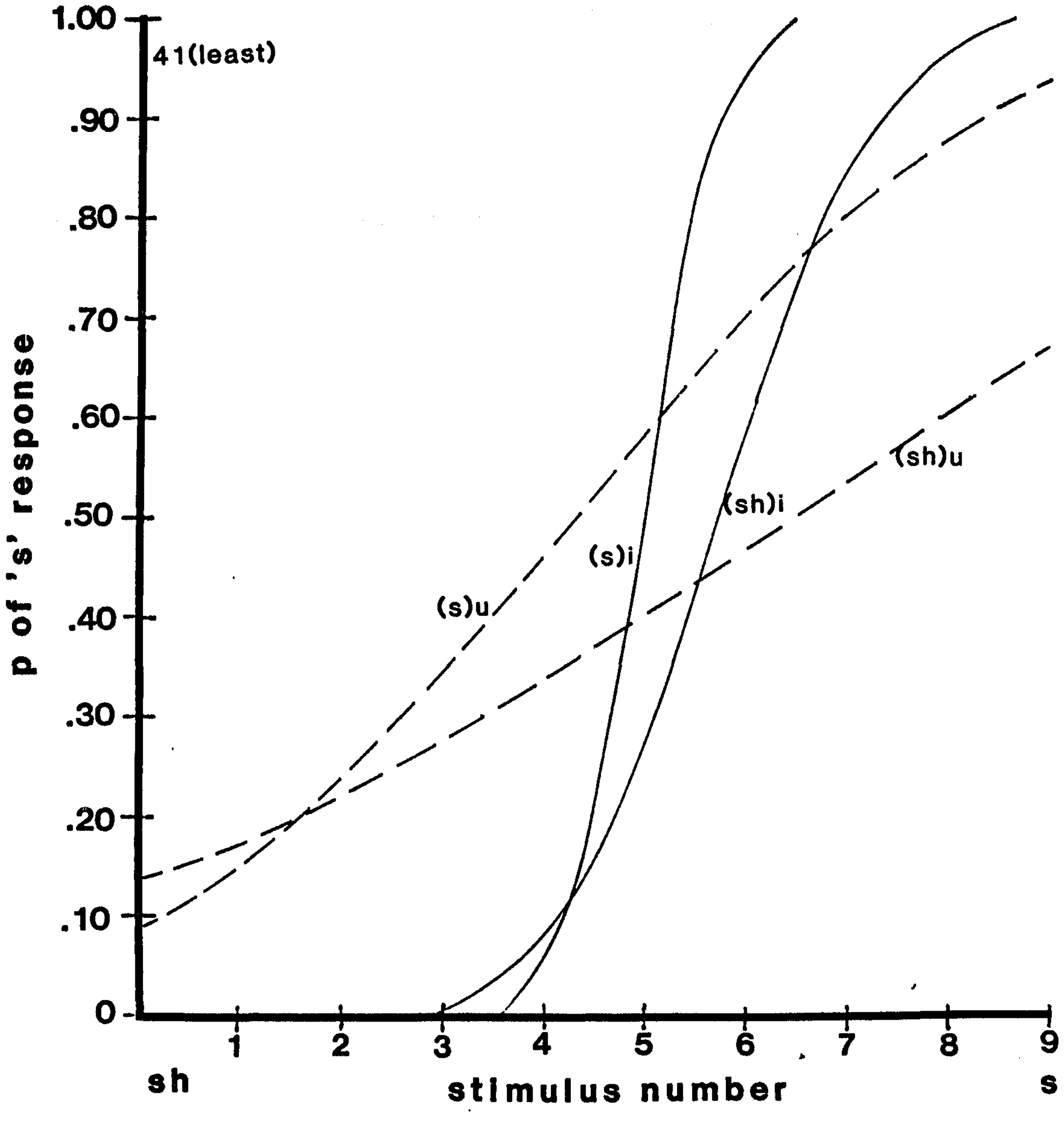


Figure 16: Identification functions for Subject 52 (most)

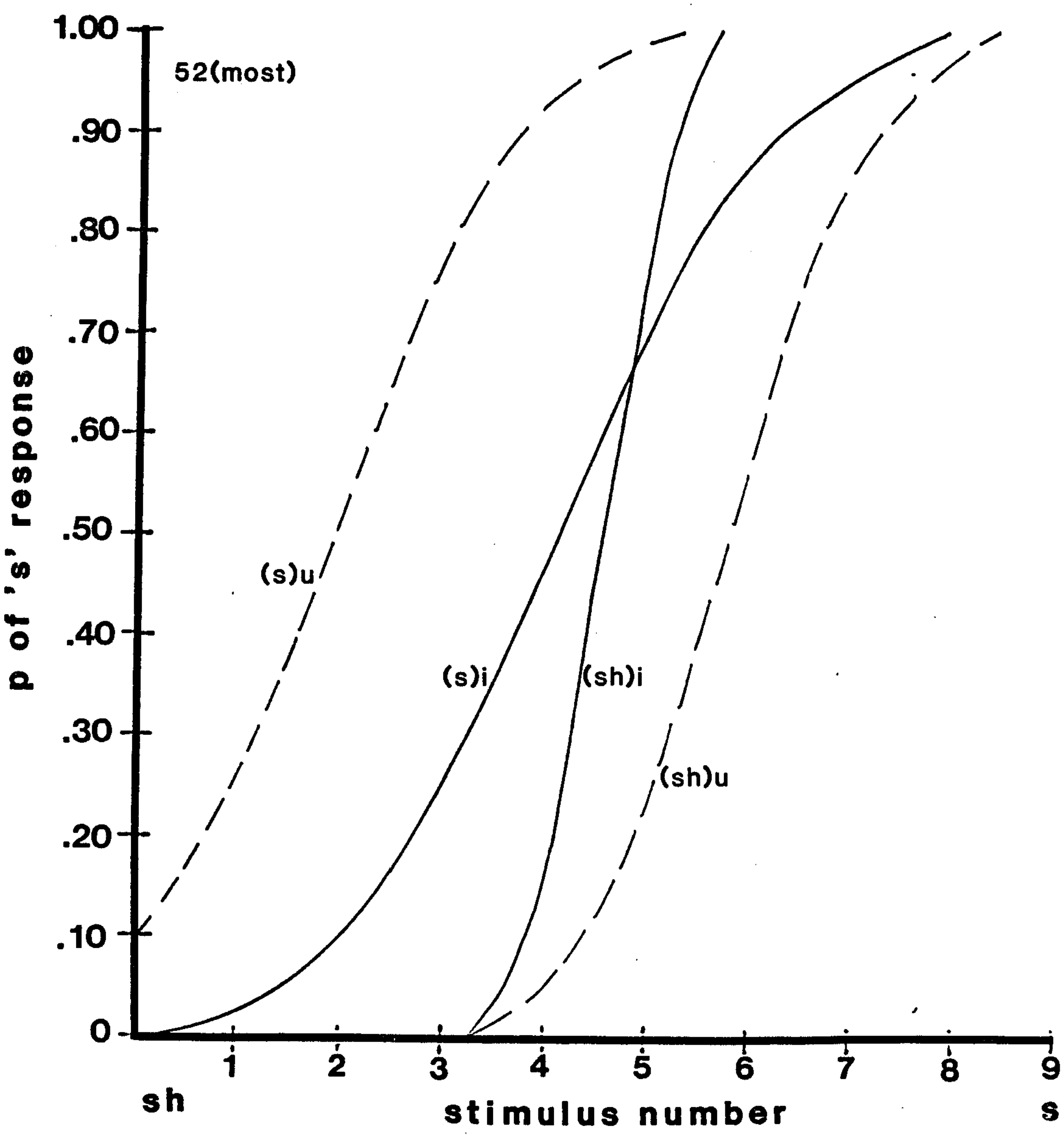


Figure 17: Identification functions for Subject 56 (least)

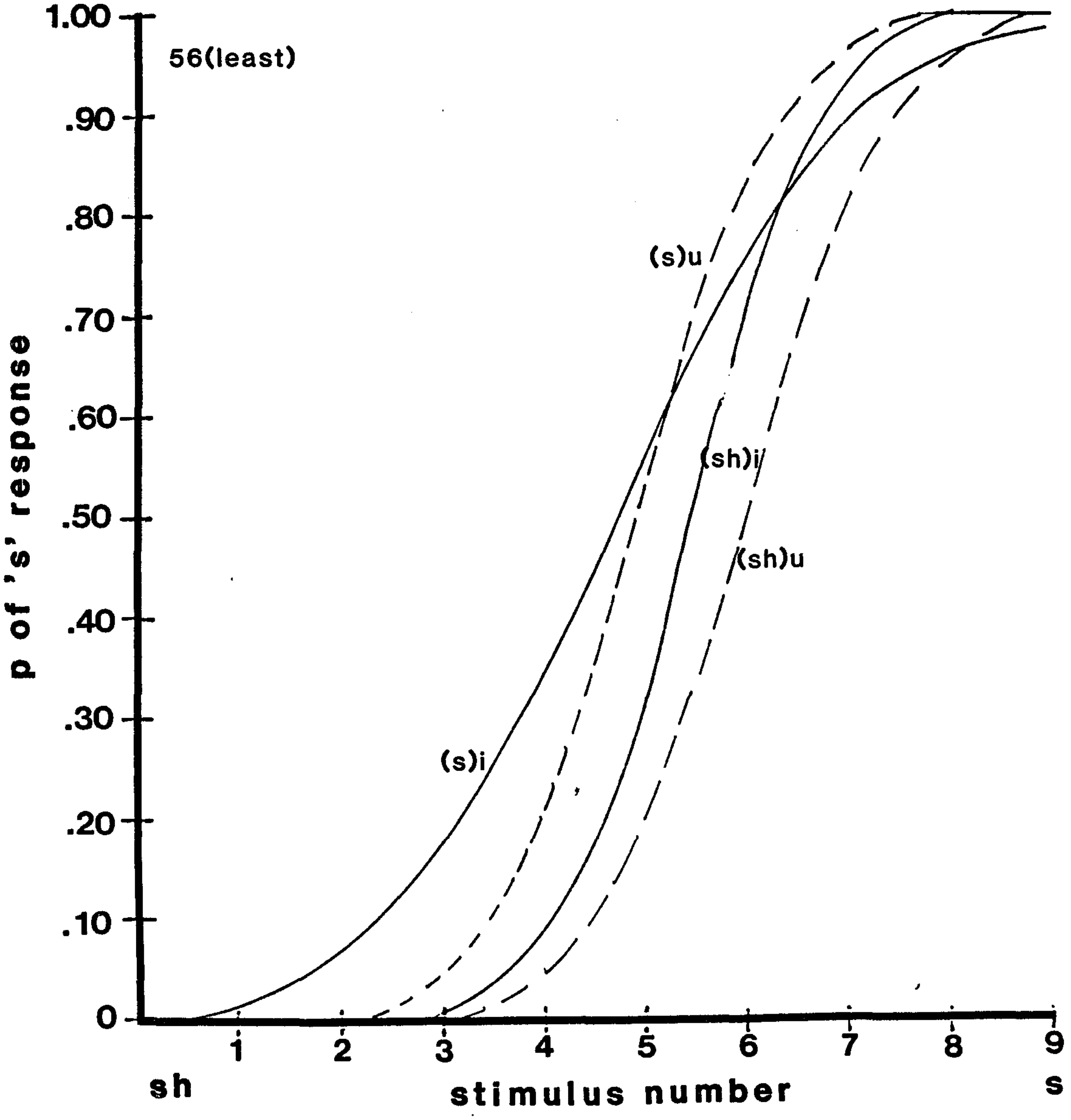


Figure 18: Identification functions for Subject 71 (most)

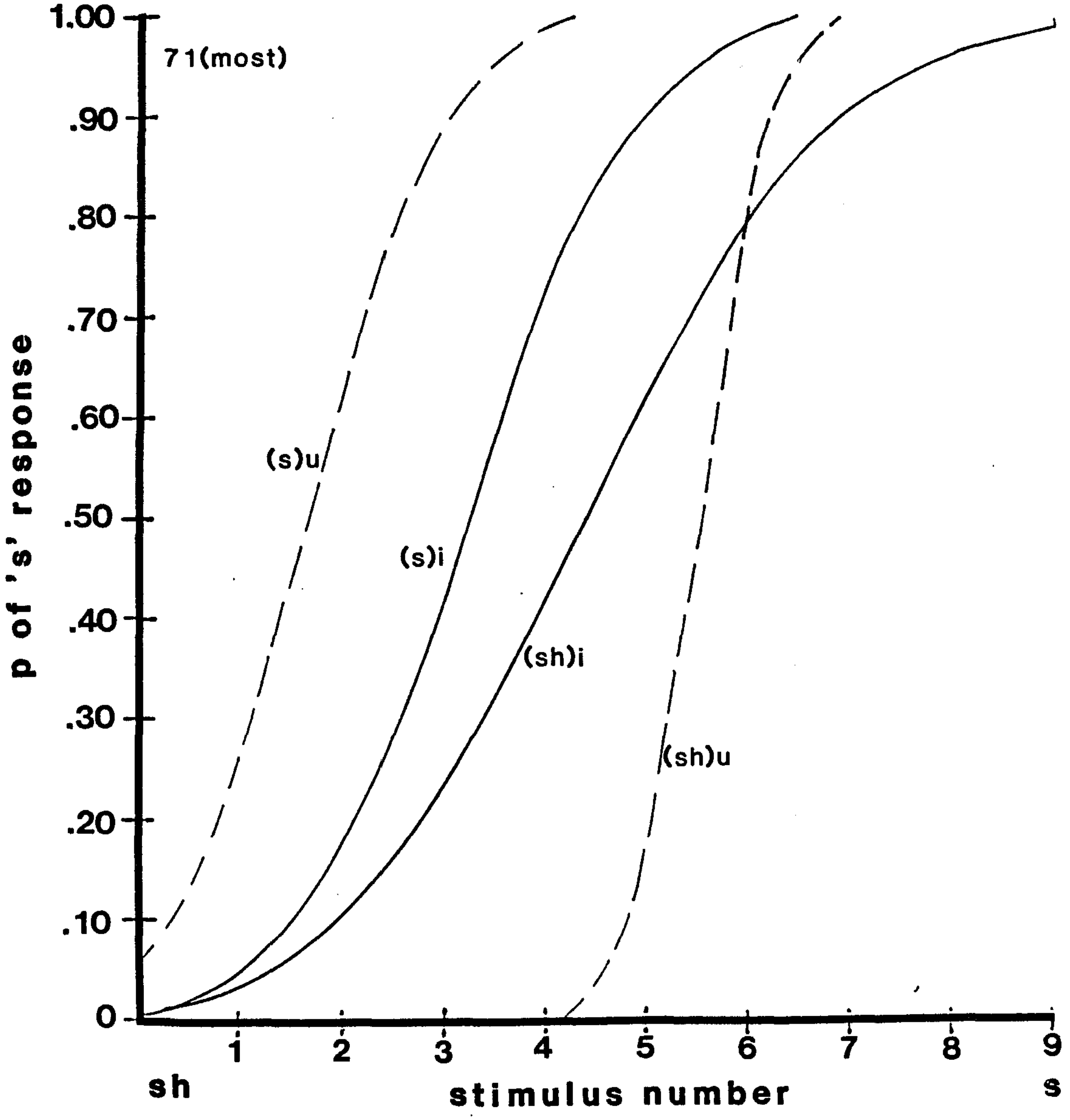


Figure 19: Identification functions for Subject 72 (least)

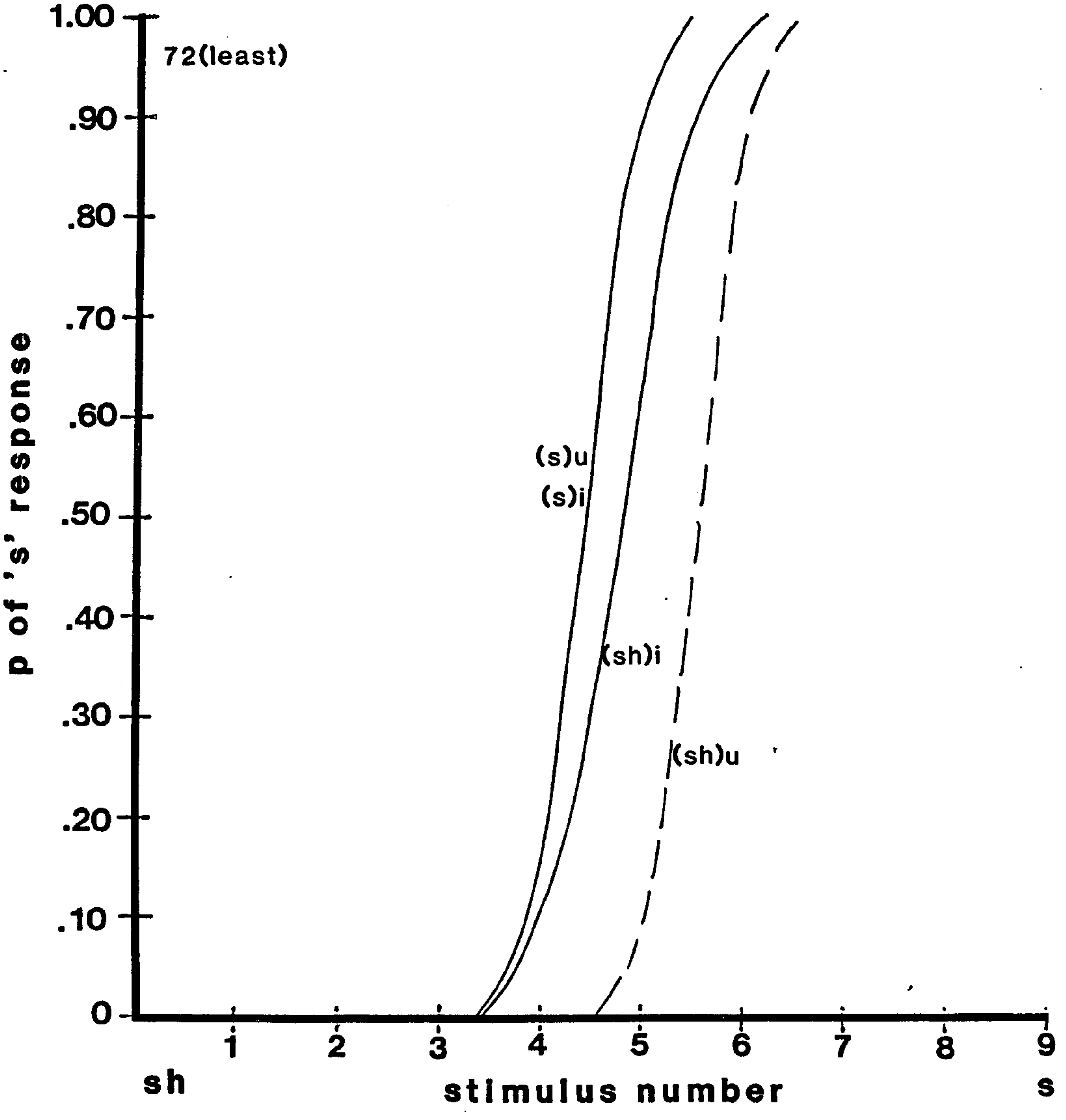
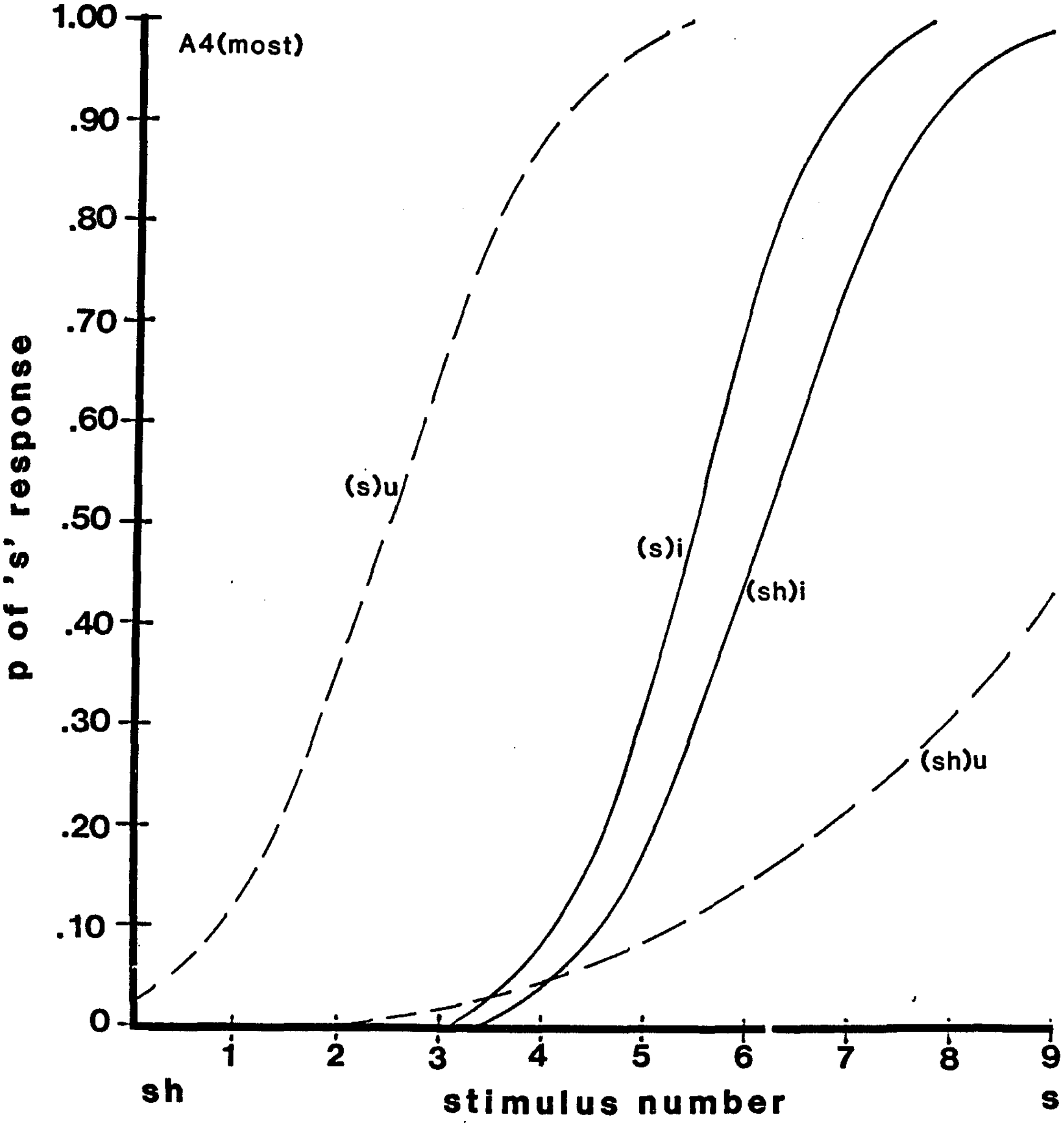


Figure 20: Identification functions for Subject A4 (most)



Materials and Equipment

The same Uher model 4200 tape recorder used for the perception experiment was used to collect production samples, with an Electrovoice model 635A microphone. The pictures used with the children in the perception experiment were also used to elicit production samples from them, and index cards with the words 'see', 'she', 'Sue', and 'shoe' printed on them were used with the adults. A Digital Equipment Corporation VAX computer and a Kay Elemetrics Digital Spectrograph were used for acoustic analyses.

Procedures

Speech samples were collected during the last testing session for each subject. The subject said the word corresponding to the picture (for children) or printed word (for adults). Pictures and printed words were presented in ten randomized blocks of four. Each word was spoken twice, as though in a two syllable word (e.g., "shoeshoe" was spoken in response to the picture of the shoe). Ten samples of each of these disyllables were collected from each subject.

The decision to have subjects say each word twice was based largely on expectations from Kent's (1983) suggestion that children coarticulate less than adults. While it may be possible to produce an initial fricative with no coarticulatory vowel effect, it seems less likely that a medial fricative could be produced without coarticulatory vowel effects. Therefore, it seemed desirable to obtain samples from initial and medial fricatives. If children failed to demonstrate coarticulatory effects for initial fricatives, it was hoped that these effects would be demonstrated for medial fricatives, thus providing some data to be compared to adults'.

Out of the 10 samples of each disyllable collected for each subject, 8 were chosen for analysis which did not contain any extraneous noises (e.g., noise due to tapping the microphone). All utterances chosen for analysis were digitized at a 20 kHz sampling rate using a low-pass filter with an upper cut-off of 9.6 kHz. Three sections were extracted from each sample used, 2 for analysis of fricative spectrum, and 1 for analysis of formant transitions.

Fricative spectra. Two sections were extracted for this analysis, one starting 100 ms before the onset of the first vowel (VO1-100 ms) and the other 35.6 ms before the onset of the second vowel (VO2-35.6 ms). Each section was 30 ms in duration, and so they extended to 70 ms before the onset of the first vowel and to 5.6 ms before the onset of the second vowel. These locations were chosen because it seemed they had the maximum likelihood of exhibiting different degrees of coarticulatory effects: The first section was further than 60 ms from the vowel, at which point Soli (1981) reported finding little evidence of vowel coarticulation, and the second section was within the region reported by Soli to demonstrate clear evidence of vowel coarticulation. Moreover, because the second section was taken from an intervocalic consonant, it could also have demonstrated evidence of perseveratory vowel effects. These sections were analyzed in 2 ways. First, Fourier analysis was done on each using a 25.6 ms Hamming window, and a centroid ("center of gravity") was computed. The centroid is a frequency average weighted by amplitude, thus reflecting the entire spectrum.

Next, the Fourier spectra of each section of each utterance were averaged and smoothed using a 200 Hz rectangular moving window. This procedure resulted in average spectra of the two sections of each of the four utterances of each subject, making a total of eight averaged spectra per subject.

Formant transitions. A 210 ms section was extracted starting 60 ms before the onset of the second vowel. Spectrographic and LPC analyses were used to resolve formant values over the last 60 ms of the fricative and first 150 ms of the vowel. From these analyses, F2 values at 30 ms before vowel onset (VO2-30 ms) and immediately after vowel onset (VO2) were recorded.

Comparison of production and perception results. Ranks for production results of adults were compared with ranks for perception results to determine whether or not there was a relation between the extent to which any individual subject coarticulated in production, and the extent to which he/she used the acoustic consequences of coarticulation in perception.

RESULTS

All tables and figures to be described in this section are provided at the end of the section because they will be referred to often. Also, note the slight differences in ordinate scales among some figures. Two-way ANOVAs (Fricative x Vowel) were done on all data collected from adults; 3-way ANOVAs (Age x Fricative x Vowel) were done on data collected from selected subjects in each age group. Thus, data for Subjects A8 and A4 were included in both sets of analyses. Results of all ANOVAs performed on production data are summarized in Table 24.

Fricative Spectra

Tables 8 and 9 and Figures 21 through 24 give mean centroids (in Hz) from both sample locations computed for each subject in each context. Looking first at the centroids for sections taken at VO1-100 ms, it is apparent that adults distinguished between /s/ and /ʃ/ samples more clearly than children did. A horizontal line could be drawn at approximately 6200 Hz on Figure 21 (adult values) to separate centroids for /s/ and /ʃ/ samples; but no such line could be drawn for the children on Figure 22. Results of the 2-way ANOVA on the adult centroids taken at

VO1-100 ms demonstrated a significant main effect of fricative only (F(1,7)=125.05, p<.01). The 3-way ANOVA on the centroids from selected subjects in each age group showed a significant main effect of fricative (F(1,5)=202.82, p<.01), and a significant interaction of Fricative x Age (F(4,5)=13.09, p=.02).

Difference scores corresponding to those computed in the perception experiment were computed for centroids, and are given in Tables 10 and 11. The first 2 columns of Table 10 indicate the extent to which centroids for samples taken starting at VO1-100 ms differed as a function of fricative. The values in parentheses indicate the percentage of change in the mean centroids (from Table 8) represented by these difference scores. For example, the 31% under "si-ʃi" for adults indicates that the mean difference score (1732 Hz) represents a 31% increase for the mean centroid value of /s(i)/ (7385 Hz) over /ʃ(i)/ (5653 Hz). As can be seen, difference scores for different fricatives produced in the same vowel context are roughly twice as large for adults as for children, and this differential effect of age is apparent even for the oldest children. That is, the 7-year-olds showed similar results to the other children.

Ratios were computed to serve as an index of the amount of difference between centroids, as a function of both fricative and vowel, and are given in Tables 12 and 13. The first 2 columns of Table 12 provide ratios describing the fricative effect on production samples taken at V01-100 ms. Again it can be seen that adults demonstrated greater differences in centroid values as a function of fricative than children, and results for all children appear similar, regardless of age.

One other descriptive statistic was computed for each subject: η^2(eta^2), the correlation ratio. This statistic is a general index of bivariate correlation that estimates the combined linear and curvilinear components of a relation. Unlike r^2, the correlation coefficient, eta^2 may be used when one variable is random and continuous, while the other is fixed and discrete, as in the present study. Eta^2 values are given by the ratio of a between groups sum of squares to a total sum of squares, and thus directly estimate the proportion of the variance in one variable (here, the centroids) that may be attributed to variation in another (here, fricative category). A separate analysis was done for each subject, and the resulting values are given in Tables 14 and 15.

Table 14 shows that 86% of the variability in adult centroids for sections taken at V01-100 ms is associated with the fricative; however, only 46% of the variability in children's centroids taken from the same location is associated with the fricative.

In general, the results described above indicate that children did not distinguish between /s/ and /ʃ/ productions to the same extent as adults did, and this age difference was apparent for children as old as 7 years. However, only results for sections taken 100 ms before the onset of the first vowel have been described. Presumably, fricative articulation at this point is as distinct as it can be, demonstrating the least effects of vowel coarticulation.

Turning to centroids derived from samples taken at VO2-35.6 ms, it can be seen from Table 9 that the distinction between centroids for different fricatives produced in the same vowel context is somewhat reduced for adults. Both difference scores and ratios indicate a reduction of approximately 7 to 9% for adults (Tables 11 and 13). This reduction in the fricative effect for adult production is apparent in Figure 23 by the fact that a horizontal line can no longer be drawn separating /s/ centroids from /ʃ/ centroids. However, a 2-way ANOVA performed on these centroids for adults still revealed a main effect of fricative ($F(1,7)=35.25$, $p<.01$). For children, there is also an overall reduction in the difference between centroid values of /s/ and /ʃ/, but it seems to be stronger in the /i/ context than in the /u/ context. Mean ratios for si/ʃi demonstrated a reduction of .06 from values at VO1-100 ms to values at VO2-35.6 ms, but mean ratios for su/ʃu are the same for both sections. It must be noted, however, that mean ratios for su/ʃu are lower than si/ʃi at both locations. Therefore, it would be misleading to say that fricatives remained as distinct approaching vowel onset as earlier during fricative production in the /u/ context for children; rather it seems that children do not distinguish well between /s/ and /ʃ/ in the /u/ context throughout the length of the fricative.

The main effect of fricative is significant for values obtained at 35.6 ms from selected subjects in each age group (F(1,5)=91.00, p<.01), but the Fricative x Age interaction is not. Thus it seems that differences between centroid values for adults as a function of fricative category have decreased enough at this point to be similar to those for children.

The eta^2 values for sections taken at VO2-35.6 ms indicate a reduction in the proportion of variability associated with fricative category for all subjects, but this reduction appears greater for children than for adults (Table 15). While this result may at first seem to contradict other measures which indicate that reduction in centroid differences between /s/ and /ʃ/ for the two sections is generally greater for adults than children, it must be remembered that children's fricatives were less distinct initially. That is, mean centroids for /s/ and /ʃ/ were closer in value at VO1-100 ms, and so a smaller reduction between means may reflect a greater increase in the areas of the sampling distributions which overlap.

In considering the effects of vowel context on fricative production, it can be seen from Table 10 and Figure 21 that vowel context had the expected effect on /s/ samples taken at VO1-100 ms for adults (centroids for /i/ tokens are higher than for /u/ tokens), but a reverse effect on /ʃ/ samples (centroids for /u/ tokens are higher than for /i/ tokens). Thus, ANOVA results of adult data taken at VO1-100 ms fail to show a significant main effect of vowel, but do show a significant interaction of Fricative x Vowel (F(1,7)=7.26, p=.03).

The mean difference in centroids between /s(i)/ and /s(u)/ at VO1-100 ms was 241 Hz for adults (Table 10). In the perception experiment, the mean difference in phoneme boundaries between /(s)i/ and /(s)u/ was 1.26 steps.

Because the step size was 200 Hz, this difference in phoneme boundaries represents a shift of approximately 250 Hz. Therefore, the mean vowel effect in production corresponds to the mean vowel effect in perception for adults.

The mean difference in phoneme boundaries between /(ʃ)i/ and /(ʃ)u/ for adults in the perception experiment was -.71 (Table 10), which represents a shift of approximately -140 Hz. The mean difference in centroids between /ʃ(i)/ and /ʃ(u)/ in production for adults was -107 Hz. Again, results from the perception experiment correspond fairly well to results from the production experiment. Furthermore, Subject A4, who demonstrated the greatest shift in phoneme boundaries between /(ʃ)i/ and /(ʃ)u/ in the perception experiment, showed the greatest shift in centroid values for /ʃ(i)/ and /ʃ(u)/. Subject A8, who demonstrated the least overall coarticulatory effects in perception, showed an effect in the opposite direction of most other adults in production: centroid values were higher for /ʃ(i)/ than for /ʃ(u)/. However, any attempt to make one-to-one correspondences between ranks of subjects in the perception experiment and ranks in the production experiment ends with these 2 subjects. For example, Subjects A5 and A2, who demonstrated weak coarticulatory effects in perception, showed strong "ʃi-ʃu" differences in production.

For children, vowel context seems to influence /s/ spectra more than for adults. For example, mean difference scores for si-su at V01-100 ms represent 6% of the mean centroid of /s(u)/ for children, but only 3% for adults (Table 10). This increased vowel effect at V01-100 ms for children is reflected in the ratios (Table 12), which are greater in value for si/su for children than adults, and eta^2 values (Table 14), which show a slightly higher proportion of variability accounted for by vowel context for children than adults. However, children's mean centroids for /ʃ(i)/ and /ʃ(u)/ are close in value at V01-100 ms, with just a slight difference in the opposite direction from adults. That is, the /ʃ(i)/ centroid is slightly higher than the /ʃ(u)/ centroid for children (Table 10).

The vowel effects on centroids demonstrated by adults and children for samples taken at V01-100 ms are generally replicated for samples taken at V02-35.6 ms (Table 11); thus it seems that any effects of perseveratory coarticulation from the first syllable were minimal. One difference is that the slight vowel effect found for children's /ʃ/ centroids at V01-100 ms is increased for children's /ʃ/ centroids at V02-35.6 ms. This increased vowel effect for children's /ʃ/ tokens taken at V02-35.6 ms is reflected in the significant main effect of vowel in the ANOVA performed on data from selected subjects (F(1,5)=22.14, p<.01). The finding that the vowel effect for children is in the direction to be expected based on vowel quality (i.e., higher centroid values before /i/ than before /u/) suggests that children were providing more information about the following vowel during fricative production. In other words, they may have started vowel articulation relatively earlier than adults did.

The mean normalized Fourier spectra are provided in Appendix E. Spectra for adults are provided first; spectra for children next. Spectra shown on the figures numbered '1' are for samples taken at V01-100 ms; spectra on figures numbered '2' are for samples taken at V02-35.6 ms. Solid lines represent spectra for samples produced in the /i/ context, and dotted lines represent spectra produced in the /u/ context.

Spectra for adults generally show consistent patterns. The /ʃ/ spectra show spectral peaks between 2000 Hz and 4000 Hz, are relatively flat out to about 8000 Hz, and then fall off. The /s/ spectra are rising, and usually do not display peaks until 5000 Hz or higher (see, for example, spectra for Subject A3). Some /s/ spectra from samples taken at VO2-35.6 ms show the beginnings of an F2 prominence at approximately 2000 Hz (again, see spectra for Subject A3). Such F2 prominences are not apparent for adults' /ʃ/ spectra at either location, or /s/ spectra at VO1-100 ms. For the most part, variation due to vowel context is minimal.

Spectra for children are not as distinct in pattern between /ʃ/ and /s/ samples as adults'. Many of the children's spectra appear flat, and any spectral peaks which can be discerned are more similar in value for the /ʃ/ and /s/ contexts (see, for example, spectra for Subject 52). More variation due to vowel context is apparent: many spectra for the /u/ context appear shifted to the left (see spectra for Subject 52). Also, F2 prominences are observed in the fricative spectra obtained at both locations for most children.

To summarize, adults generally distinguished between /s/ and /ʃ/ more clearly in production than children did, and children seem to have displayed more evidence of vowel identity in the fricative. The distinctive quality of fricatives was reduced somewhat as all speakers approached vowel onset. Also, children showed an increase in the amount of information concerning vowel identity which was available in /ʃ/ tokens as they got closer to vowel onset.

The patterns of results described here for fricative spectra seem to separate adults from children of all ages. Results for children do not appear to vary systematically as a function of age, and results for all subjects show little systematic variation as a function of the extent to which they demonstrated sensitivity to coarticulatory effects in perception.

Second Formant Values

Second formant values could not be obtained at VO2-30 ms from the LPC analysis for 3 of the 4 adult males; the lowest formant values available for these subjects were F3 values. Inspection of the spectra and sample spectrograms (Appendix F) reveals the reason: There are no spectral prominences in the F2 region for many of the adults; however, F2 prominences are observable in the spectra of most children. In addition, the spectrograms show that there is energy in the F2 region within the fricative portion of most syllables for children, but not for adults. This trend is particularly apparent for /s/ tokens (compare, for example, spectrograms for Subjects A4 and 71).

The major difference between F2 values taken at VO2-30 ms and VO2 is that differences in adults' F2 values at VO2-30 ms (for the adults whose F2 values could be measured) as a function of both fricative and vowel were just barely significant at the .05 level, but differences in F2 values at VO2 were highly significant as a function of fricative and vowel (Table 24). Children showed highly significant effects of both fricative and vowel at both locations. As with centroid values, F2 values indicate that children provided more vowel information within the fricative portion of the syllable. Because F2 values obtained at VO2-30 ms and VO2 reveal similar patterns, and because values were available for all subjects at VO2, the remainder of this discussion will focus on F2 values obtained at VO2.

Tables 17, 19, and 21 indicate that second formant values are generally higher in syllables produced with /ʃ/ than syllables produced with /s/. This finding is also observed in Figures 27 and 28 by the fact that circles, representing /ʃ/, are generally higher than squares, representing /s/. In other words, there is a fricative effect on F2 values at vowel onset. Three exceptions to this pattern are found: both 3-year-old subjects and Subject 56 demonstrated higher F2 values in /su/ than in

/ʃu/. Overall, younger children exhibited a smaller effect of fricative on F2 in the /u/ context than adults (Table 19). This age effect is apparent in ratios as well: Adults showed a mean value of 1.14 for ʃu/su, and children showed a mean value of 1.08 (Table 21). In the /i/ context, however, children demonstrated a fricative effect similar in size to that of adults. Eta^2 results indicate that F2 values at VO2 were associated with fricative category less strongly for children than adults (Table 23).

Turning to the influence of vowel context on F2, it can be seen that children's values were more strongly associated with the following vowel quality than were adults'. This age difference is apparent in all measures: (1) Difference scores for vowel effects are greater for children than for adults (Tables 18 and 19); (2) Ratios indicating vowel effects are greater for children than for adults (Tables 20 and 21); and (3) The proportion of variability in F2 associated with vowel category is greater for children than for adults (Tables 22 and 23). This enhanced vowel effect shown by children is also apparent in Figure 28 by the fact that open symbols, representing /i/, are generally higher than filled symbols, representing /u/. In general, it seems that children are closer to the steady-state vowel at voicing onset than adults.

The patterns of results described above are all apparent in the ANOVA results. The 2-way ANOVA (Fricative x Vowel) performed on F2 values obtained for adults at VO2 reveals highly significant main effects of fricative and vowel, as well as a significant interaction for Fricative x Vowel (Table 24). This significant interaction term for adults reflects the greater vowel effect on samples with /s/ as compared to samples with /ʃ/. For example, the mean adult ratio for F2 values obtained at VO2 is 1.14 for si/su, and 1.07 for ʃi/ʃu (Table 21).

The 3-way ANOVA (Age x Fricative x Vowel) performed on data for selected subjects shows significant main effects of both fricative and vowel, as the adult analysis did. The main effect of age was just barely significant for F2 values obtained at VO2 (F(4,5)=5.45, p=.05). Actually, this effect probably indicates only that younger (and presumably smaller) subjects demonstrated higher F2 values than older subjects.

Sample spectrograms, shown in Appendix F, are for acoustic segments taken starting at VO2-60 ms and ending at VO2+150 ms. Inspection of these spectrograms shows that children generally demonstrate flatter vocalic transitions than adults (cf. Kent, 1983). That is, there is less movement after vowel onset in children's samples (see, for example, spectrograms for Subject 35). In addition, children's spectrograms show more evidence of noise than adults.

Summarizing F2 measurements, children demonstrated greater differences in F2 values within the fricative portion of the syllable as a function of the following vowel quality than adults demonstrated; however, children showed smaller differences in F2 values at vowel onset as a function of the preceding fricative than did adults.

The Relation Between Perception and Production

Adults were ranked on production results according to the degree of coarticulatory effects demonstrated. Two measures were used for assigning ranks: (1) differences in centroid values for si-su at VO1-100 ms; and (2) differences in F2 at VO2 for ʃu/su. These measures seemed to correspond most closely to those used to assign ranks in the perception experiment. That is, one measure seemed clearly to reflect a vowel effect on fricative production (si-su), and the other seemed to reflect a transition effect on vowel production (ʃu-su). Ranks assigned to adults for production results are given in Table 25.

Numbers in parentheses are ranks for perception results. Spearman's rank order correlation coefficient between perception and production has a nonsignificant value of -.02. Thus, there appears to be no correlation in the size of coarticulatory effects demonstrated in perception and production for individual subjects. This lack of correlation for individual subjects appears to hold for children as well: Inspection of production data shows no systematic relation for children who demonstrated the greatest and least effects in perception.

Table 8: Mean centroids (in Hz) for samples taken starting at 100 ms before onset of first vowel (VO1-100 ms)

	ʃ(i)	ʃ(u)	s(i)	s(u)
Adults				
A4(most)	5017	5331	6916	7056
*A1(2)	5774	5747	7636	6487
A6(3)	5543	5533	6863	6385
*A3(4)	5852	5920	8006	8041
A7(5)	5575	5685	6861	6989
A5(6)	5814	6074	7161	7022
*A2(7)	5935	6179	7898	7732
*A8(least)	5714	5610	7740	7675
$^{2}\bar{X}$=	5653	5760	7385	7174
males=	5487	5656	6950	6863
females=	5819	5864	7820	7484
Children				
71(most)	5901	5761	6708	6666
72(least)	5929	5835	6708	6550
52(most)	6014	6122	7262	6672
56(least)	5802	6149	6932	6380
44(most)	6841	6600	7359	6942
41(least)	5896	5987	6900	5949
32(most)	6516	7081	7441	7424
35(least)	5971	5849	7272	7013
$\bar{X}$=	6109	6173	7073	6700

*female adults

Table 9: Mean centroids (in Hz) for samples taken starting at 35.6 ms before onset of second vowel (VO2-35.6 ms)

	ʃ(i)	ʃ(u)	s(i)	s(u)
Adults				
A4(most)	5041	4952	5573	5820
*A1(2)	5108	5052	6735	5974
A6(3)	5077	4959	5823	5779
*A3(4)	5657	5941	7562	7519
A7(5)	5486	5852	6431	6436
A5(6)	5663	6076	6633	6553
*A2(7)	5314	5403	7804	7176
*A8(least)	5260	5145	6210	6081
X̄=	5326	5423	6597	6418
males=	5317	5460	6115	6147
females=	5335	5385	7078	6688
Children				
71(most)	5804	5314	6170	6454
72(least)	5259	5207	5832	5577
52(most)	5432	5243	6176	5791
56(least)	5639	5846	6573	5884
44(most)	6342	6224	6851	6658
41(least)	5544	5713	5907	5423
32(most)	5957	5839	6562	6528
35(least)	5576	5036	5888	5793
X̄=	5694	5553	6245	6014

Table 10: Difference scores (in Hz) for mean centroids taken starting at V01-100 ms

	Fricative Effect		Vowel Effect	
Adults	si-ʃi	su-ʃu	ʃi-ʃu	si-su
A4(most)	1898	1725	-314	-140
*A1(2)	1862	740	27	1149
A6(3)	1320	852	10	478
*A3(4)	2154	2121	-68	-35
A7(5)	1286	1304	-110	-128
A5(6)	1347	948	-260	139
*A2(7)	1963	1553	-244	166
*A8(least)	2026	2065	104	65
$\bar{X}$=	1732	1414	-107	241
	(31%)[a]	(25%)	(-2%)	(3%)
Children				
71(most)	807	905	140	42
72(least)	779	715	94	158
52(most)	1248	550	-108	590
56(least)	1130	1361	-347	552
44(most)	518	342	241	417
41(least)	1004	-38	-91	951
32(most)	925	343	-565	17
35(least)	1301	1164	122	259
$\bar{X}$=	964	668	23	373
	(16%)	(11%)	(0%)	(6%)

[a] percentage of change in mean centroid value

Table 11: Difference scores (in Hz) for mean centroids of samples taken at VO2-35.6 ms

	Fricative Effect		Vowel Effect	
Adults	si-ʃi	su-ʃu	ʃi-ʃu	si-su
A4(most)	532	868	89	
247				
*A1(2)	1627	922	56	761
A6(3)	746	820	118	44
*A3(4)	1905	1578	-284	43
A7(5)	945	584	-366	-5
A5(6)	970	477	-413	80
*A2(7)	2490	1773	-89	628
*A8(least)	950	936	115	129
$\bar{X}$=	1271	995	-97	179
	(24%)	(18%)	(-2%)	(3%)
Children				
71(most)	366	1140	490	-284
72(least)	573	370	52	255
52(most)	744	548	189	599
56(least)	934	38	-207	689
44(most)	509	434	118	193
41(least)	363	-290	-169	484
32(most)	605	689	118	34
35(least)	312	757	540	95
$\bar{X}$=	551	416	141	258
	(10%)	(7%)	(3%)	(4%)

<u>Table 12</u>: Ratios of mean centroids of samples taken starting at VO1-100 ms

	Fricative Effect		Vowel Effect	
Adults	si/ʃi	su/ʃu	ʃi/ʃu	si/su
A4(most)	1.38	1.32	.94	.98
*A1(2)	1.32	1.13	1.00	1.18
A6(3)	1.24	1.15	1.00	1.07
*A3(4)	1.37	1.38	.99	1.00
A7(5)	1.23	1.23	.98	.98
A5(6)	1.23	1.16	.96	1.02
*A2(7)	1.37	1.36	.99	1.00
*A8(least)	1.35	1.37	1.02	1.01
$\bar{X}$=	1.31	1.26	.99	1.03
Children				
71(most)	1.14	1.16	.98	1.01
72(least)	1.13	1.12	1.02	1.02
52(most)	1.21	1.09	.98	1.09
56(least)	1.19	1.04	.94	1.09
44(most)	1.08	1.05	1.04	1.06
41(least)	1.17	.99	.98	1.16
32(most)	1.14	1.05	.92	1.00
35(least)	1.22	1.20	1.02	1.04
$\bar{X}$=	1.16	1.09	.99	1.06

Table 13: Ratios of mean centroids of samples taken starting at VO2-35.6 ms

	Fricative Effect		Vowel Effect	
Adults	si/ʃi	su/ʃu	ʃi/ʃu	si/su
A4(most)	1.11	1.18	1.02	.96
*A1(2)	1.18	1.18	1.13	1.13
A6(3)	1.15	1.17	1.02	1.01
*A3(4)	1.34	1.27	.95	1.01
A7(5)	1.17	1.01	.94	1.00
A5(6)	1.17	1.08	.93	1.01
*A2(7)	1.47	1.33	.98	1.09
*A8(least)	1.18	1.18	1.02	1.02
$\bar{X}$=	1.22	1.18	1.00	1.03
Children				
71(most)	1.06	1.21	1.09	.96
72(least)	1.11	1.07	1.01	1.05
52(most)	1.14	1.10	1.04	1.07
56(least)	1.17	1.01	1.00	1.12
44(most)	1.08	1.07	1.02	1.03
41(least)	1.07	.95	.97	1.09
32(most)	1.10	1.12	1.02	1.01
35(least)	1.06	1.15	1.11	1.02
$\bar{X}$=	1.10	1.09	1.03	1.04

Table 14: Eta^2 for mean centroids of samples taken starting at VO1-100 ms

The proportion of variability in the mean centroid associated with the:

	fricative	vowel
Adults		
A4(most)	.88	.01
*A1(2)	.78	.07
A6(3)	.65	.03
*A3(4)	.95	.00
A7(5)	.85	.01
A5(6)	.85	.00
*A2(7)	.92	.00
*A8(least)	.97	.00
$\bar{X}$=	.86	.02
Children		
71(most)	.59	.01
72(least)	.58	.02
52(most)	.55	.04
56(least)	.38	.01
44(most)	.31	.18
41(least)	.12	.09
32(most)	.34	.06
35(least)	.84	.14
$\bar{X}$=	.46	.07

Table 15: Eta^2 for mean centroids taken starting at VO2-35.6 ms

The proportion of variability in the mean centroid associated with the:

	fricative	vowel
Adults		
A4(most)	.65	.01
*A1(2)	.76	.08
A6(3)	.61	.01
*A3(4)	.90	.00
A7(5)	.63	.04
A5(6)	.56	.03
*A2(7)	.90	.01
*A8(least)	.72	.01
$\bar{X}$=	.72	.02
Children		
71(most)	.26	.00
72(least)	.22	.02
52(most)	.28	.06
56(least)	.27	.07
44(most)	.43	.05
41(least)	.00	.02
32(most)	.17	.00
35(least)	.29	.10
$\bar{X}$=	.24	.04

Table 16: Mean F2 values at 30 ms before onset of the second vowel (VO2-30 ms)

Adults	ʃi	ʃu	si	su
A4(most)	----	----	----	----
*A1(2)	2332	2106	2156	1876
A6(3)	1812	1698	1676	1523
*A3(4)	2377	2384	2425	2265
A7(5)	----	----	----	----
A5(6)	----	----	----	----
*A2(7)	2111	2004	2052	1957
*A8(least)	2303	2141	2088	2126
males(1)=	1812	1698	1676	1523
females(4)=	2281	2159	2180	2056
Children				
71(most)	2609	2524	2433	2198
72(least)	2636	2588	2325	2017
52(most)	2749	2370	2776	2301
56(least)	3224	2695	3090	2557
44(most)	3100	2758	2589	2285
41(least)	2884	2540	2554	2429
32(most)	2899	2383	2545	2272
35(least)	2884	2380	2605	2438
$\bar{X}$=	2873	2530	2646	2266

Table 17: Mean F2 values taken immediately after onset of second vowel (VO2)

	ʃi	ʃu	si	su
Adults				
A4(most)	1796	1783	1709	1587
*A1(2)	2346	2064	2162	1773
A6(3)	1765	1604	1613	1411
*A3(4)	2377	2253	2242	2109
A7(5)	2063	1924	1949	1549
A5(6)	1968	1913	1746	1572
*A2(7)	2015	1884	1977	1753
*A8(least)	2356	2184	2200	1925
males=	1898	1806	1754	1530
females=	2274	2096	2145	1890
Children				
71(most)	2596	2468	2472	2178
72(least)	2571	2460	2304	1964
52(most)	2793	2336	2700	2249
56(least)	3232	2449	3039	2506
44(most)	3000	2772	2652	2289
41(least)	2951	2434	2670	2278
32(most)	2925	2189	2565	2200
35(least)	2827	2374	2764	2460
$\bar{X}$=	2862	2435	2646	2266

Table 18: Difference scores for F2 values at V02-30 ms

	Transition Effect		Vowel Effect	
Adults	ʃi-si	ʃu-su	ʃi-ʃu	si-su
A4(most)	----	----	----	----
*A1(2)	176	230	226	280
A6(3)	136	175	114	153
*A3(4)	-48	119	-7	160
A7(5)	----	----	----	----
A5(6)	----	----	----	----
*A2(7)	59	47	107	95
*A8(least)	215	15	162	-38
males(1)=	136(8%)	175(11%)	114(7%)	153(10%)
females(4)=	101(5%)	103(5%)	122(6%)	124(6%)
Children				
71(most)	176	326	85	235
72(least)	311	571	48	308
52(most)	-27	69	379	475
56(least)	134	138	529	533
44(most)	511	473	342	304
41(least)	330	111	344	125
32(most)	354	111	516	273
35(least)	279	-58	504	167
$\bar{X}$=	259(10%)	218(10%)	343(14%)	303(14%)

Table 19: Difference scores (in Hz) for F2 at VO2

	Fricative Effect		Vowel Effect	
Adults	ʃi-si	ʃu-su	ʃi-ʃu	si-su
A4(most)	87	196	13	122
*A1(2)	184	291	300	389
A6(3)	152	193	161	202
*A3(4)	135	144	135	133
A7(5)	114	375	139	400
A5(6)	222	341	55	174
*A2(7)	38	131	131	224
*A8(most)	172	259	172	275
males=	144(8%)	276(18%)	92(5%)	225(15%)
females=	132(6%)	206(11%)	185(9%)	255(13%)
Children				
71(most)	124	290	128	294
72(least)	267	496	111	340
52(most)	93	87	457	451
56(least)	193	-57	783	533
44(most)	348	483	228	363
41(least)	281	156	517	392
32(most)	360	-11	736	365
35(least)	63	-86	453	304
$\check{X}$=	216(8%)	170(8%)	427(18%)	380(17%)

Table 20: Ratios of F2 values at VO2-30 ms

	Fricative Effect		Vowel Effect	
Adults	ʃi/si	ʃu/su	ʃi/ʃu	si/su
A4(most)	----	----	----	----
*A1(2)	1.08	1.12	1.11	1.15
A6(3)	1.08	1.11	1.07	1.10
*A3(4)	.98	1.05	1.00	1.07
A7(5)	----	----	----	----
A5(6)	----	----	----	----
*A2(7)	1.03	1.02	1.05	1.05
*A8(least)	1.10	1.01	1.08	.98
$\bar{X}$=	1.05	1.06	1.06	1.07
Children				
71(most)	1.07	1.15	1.03	1.11
72(least)	1.13	1.28	1.02	1.15
52(most)	.99	1.03	1.16	1.21
56(least)	1.04	1.05	1.20	1.21
44(most)	1.20	1.21	1.12	1.13
41(least)	1.13	1.05	1.14	1.05
32(most)	1.14	1.05	1.22	1.12
35(least)	1.11	.98	1.21	1.07
$\bar{X}$=	1.10	1.10	1.14	1.13

Table 21: Ratios of F2 values taken at VO2

	Fricative Effect		Vowel Effect	
Adults	ʃi/si	ʃu/su	ʃi/ʃu	si/su
A4(most)	1.05	1.12	1.01	1.08
*A1(2)	1.09	1.16	1.14	1.22
A6(3)	1.09	1.14	1.10	1.14
*A3(4)	1.06	1.07	1.06	1.06
A7(5)	1.06	1.24	1.07	1.26
A5(6)	1.13	1.22	1.03	1.11
*A2(7)	1.02	1.07	1.07	1.13
*A8(least)	1.07	1.13	1.08	1.14
$\bar{X}$=	1.07	1.14	1.07	1.14
Children				
71(most)	1.05	1.13	1.05	1.13
72(least)	1.12	1.25	1.05	1.17
52(most)	1.03	1.04	1.20	1.20
56(least)	1.06	.98	1.32	1.21
44(most)	1.13	1.21	1.08	1.16
41(least)	1.09	1.07	1.21	1.19
32(most)	1.14	1.00	1.34	1.17
35(least)	1.04	.97	1.21	1.12
$\bar{X}$=	1.08	1.08	1.18	1.17

Table 22: Eta^2 for F2 values at VO2-30 ms
The proportion of variability in F2 associated with the:

	fricative	vowel
Adults		
A4(most)	---	---
*A1(2)	.31	.49
A6(3)	.34	.25
*A3(4)	.02	.10
A7(5)	---	---
A5(6)	---	---
*A2(7)	.07	.24
*A8(least)	.30	.09
$\bar{X}$=	.21	.23
Children		
71(most)	.37	.15
72(least)	.69	.11
52(most)	.00	.61
56(least)	.05	.76
44(most)	.64	.28
41(least)	.30	.33
32(most)	.15	.44
35(least)	.06	.57
$\bar{X}$=	.28	.41

Table 23: Eta^2 for F2 values taken at VO2

The proportion of variability in F2 associated with:

	fricative	vowel
Adults		
A4(most)	.45	.10
*A1(2)	.29	.58
A6(3)	.38	.42
*A3(4)	.33	.29
A7(5)	.29	.35
A5(6)	.67	.11
*A2(7)	.13	.59
*A8(least)	.40	.46
$\bar{X}$=	.37	.36
Children		
71(most)	.22	.23
72(least)	.52	.18
52(most)	.03	.69
56(least)	.01	.73
44(most)	.60	.30
41(least)	.09	.47
32(most)	.07	.67
35(least)	.00	.65
$\bar{X}$=	.19	.49

Table 24: Summary of analyses of variance performed on production data[a]

	centroids		F2	
	VO1-100 ms	VO2-35.6 ms	VO2-30 ms	VO2
ADULTS ONLY				
fricative	<.01	<.01	=.04	<.01
vowel	NS	NS	=.05	<.01
fric x vowel	=.03	NS	NS	<.01
MOST & LEAST OF EACH AGE GROUP				
age	NS	NS	=.02[b]	=.05
fricative	<.01	<.01	<.01	<.01
vowel	NS	<.01	<.01	<.01
age x fric	<.01	NS	NS	NS
age x vowel	NS	NS	=.03	NS
fric x vowel	=.02	NS	NS	NS
fric x vowel x age	NS	NS	NS	NS

[a] mean values in Hz used for all analyses; analyses done separately for adults, and subjects in each age group identified as demonstrating greatest and least coarticulatory effects in perception experiment

[b] because F2 values at VO2-30 ms were not available for Subject A4, the adult demonstrating the greatest coarticulatory effects in perception, values obtained from the subject with the next highest ranking were used for this analysis (Subject A1)

Table 25: Ranks assigned to adults for productions results

Subject	Production Rank
A4(most)[a]	8
*A1(2)	1
A6(3)	3
*A3(4)	7
A7(5)	4
A5(6)	2
*A2(7)	6
*A8(least)	5

[a] Ranks in parentheses are for perception results

Figure 21: Mean centroids for adults at VO1-100 ms

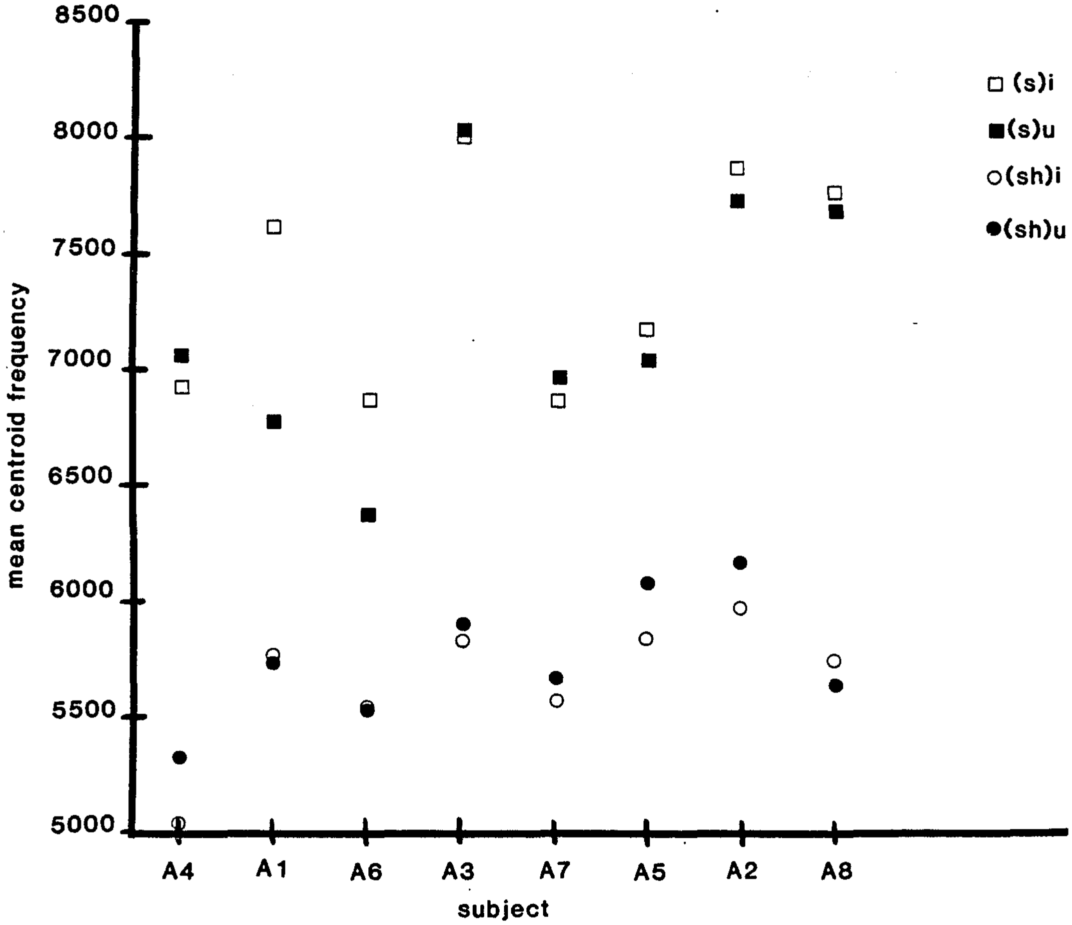


Figure 22: Mean centroids for selected subjects at V01-100 ms

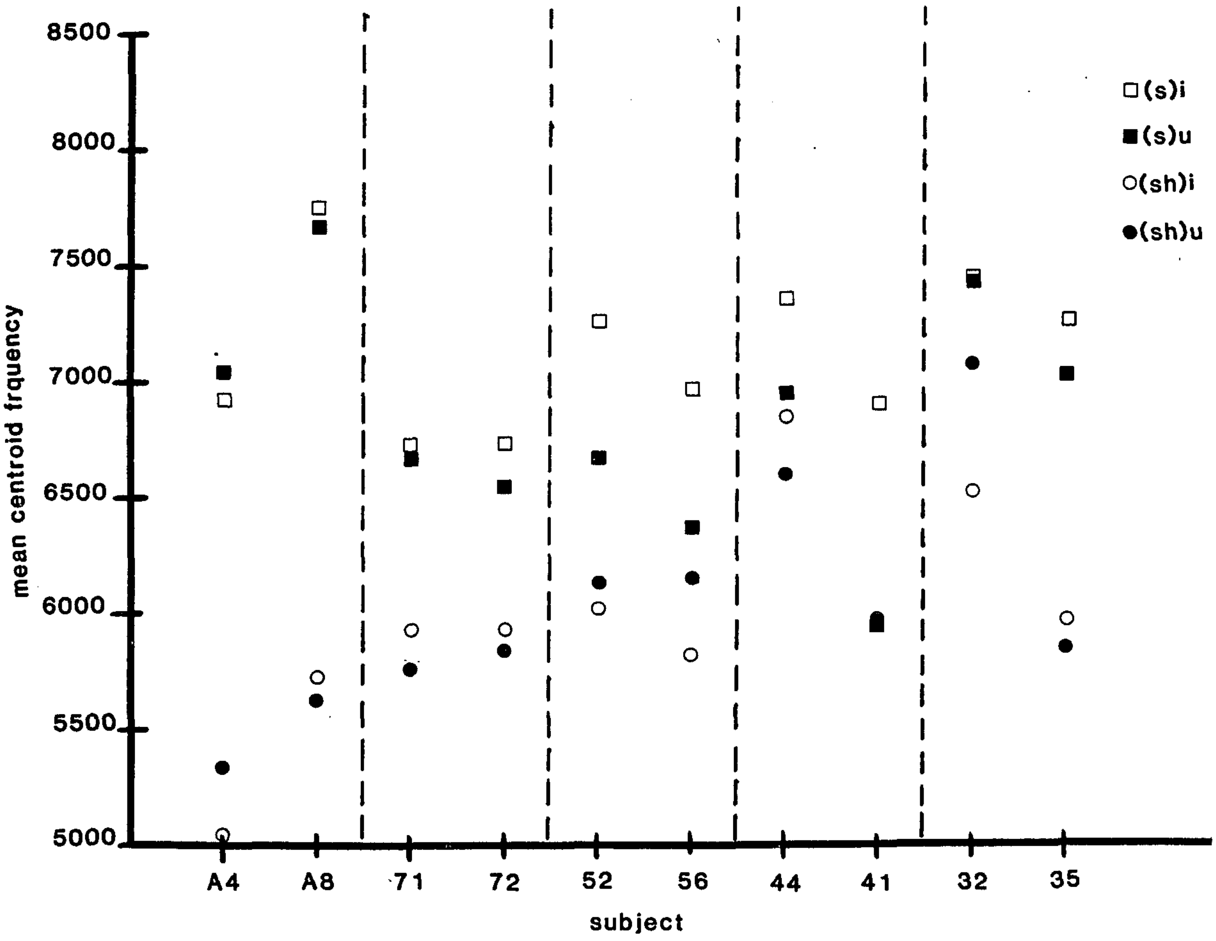


Figure 23: Mean centroids for adults at VO2-35.6 ms

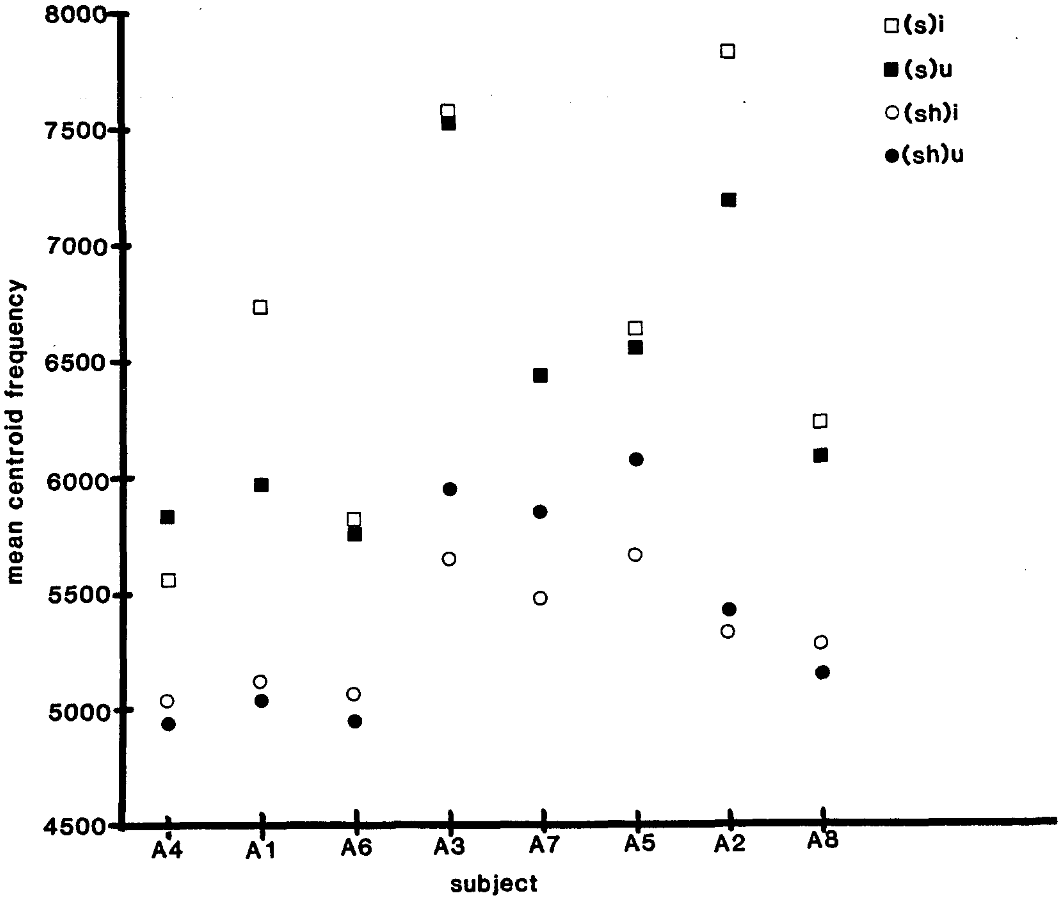


Figure 24: Mean centroids for selected subjects at VO2-35.6 ms

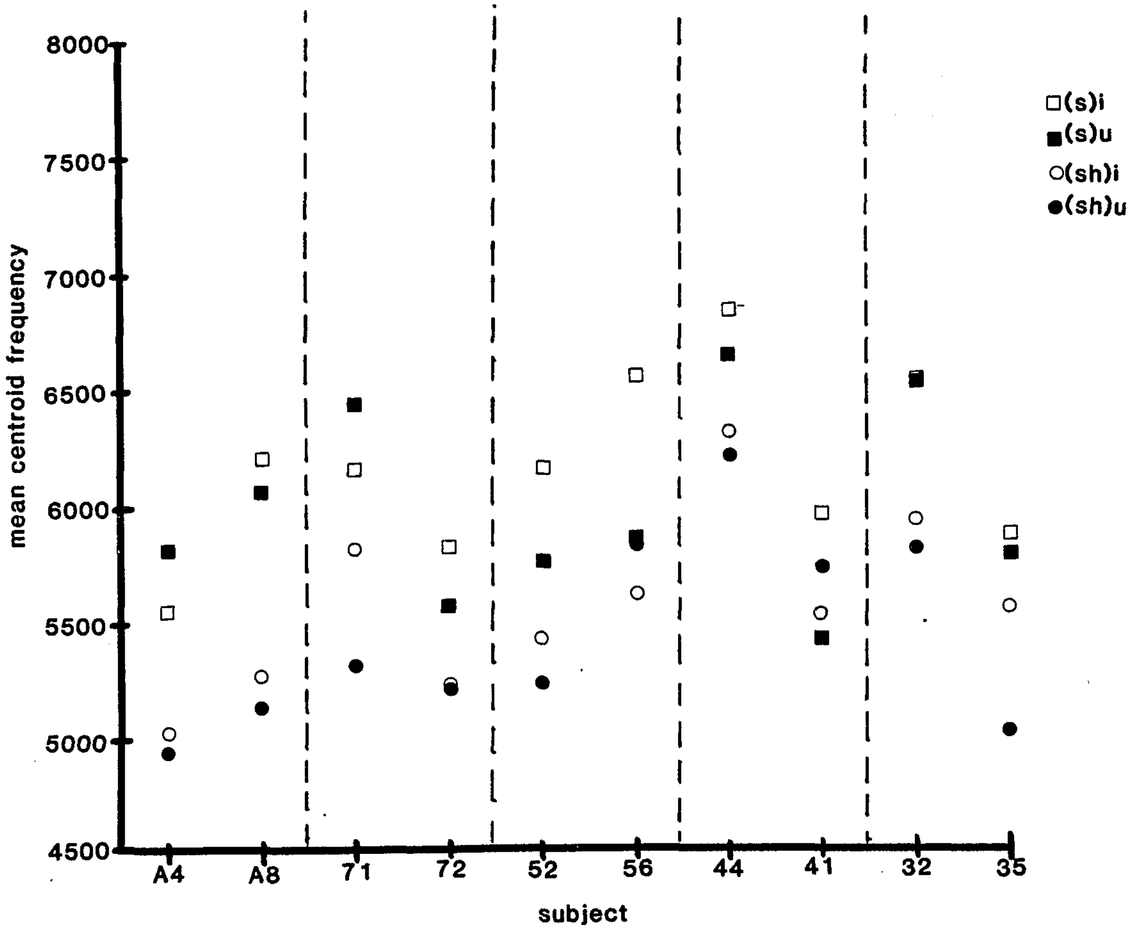


Figure 25: Mean F2 values for adults at VO2-30 ms

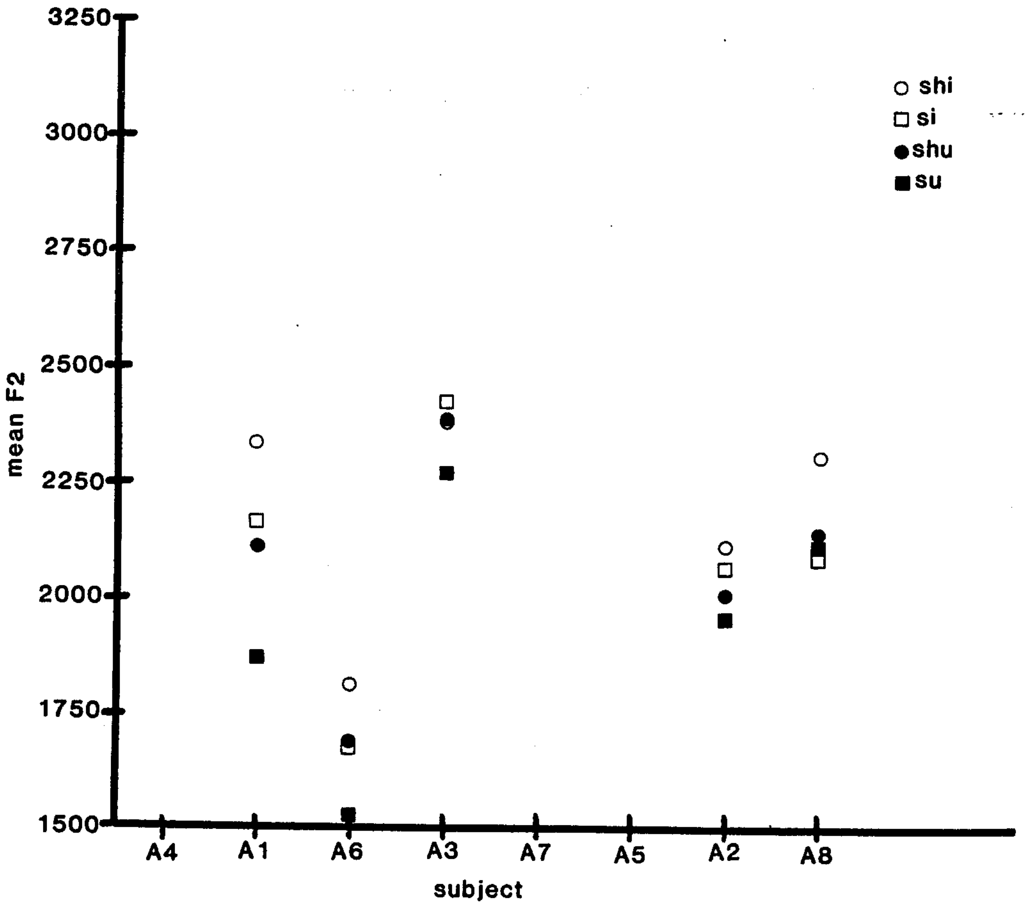


Figure 26: Mean F2 values for selected subjects at VO2-30 ms

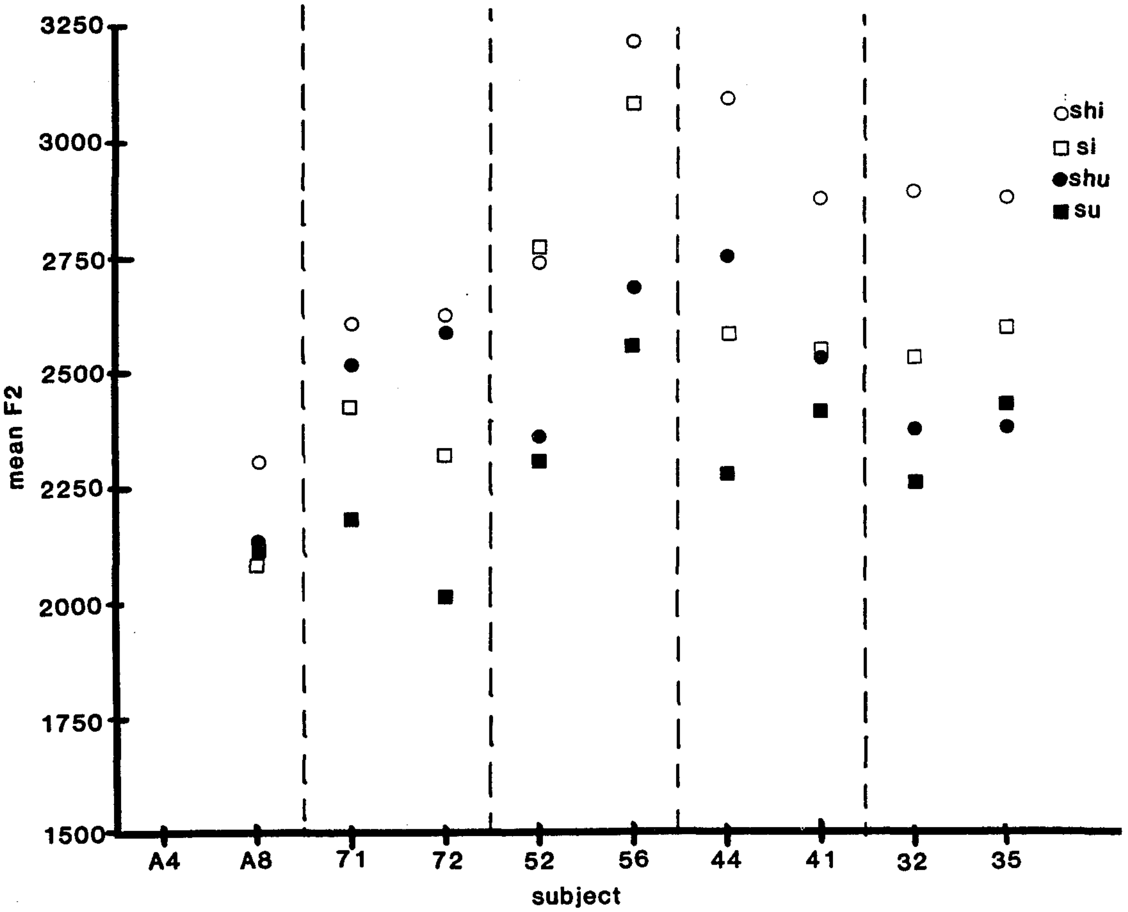


Figure 27: Mean F2 values for adults at V02

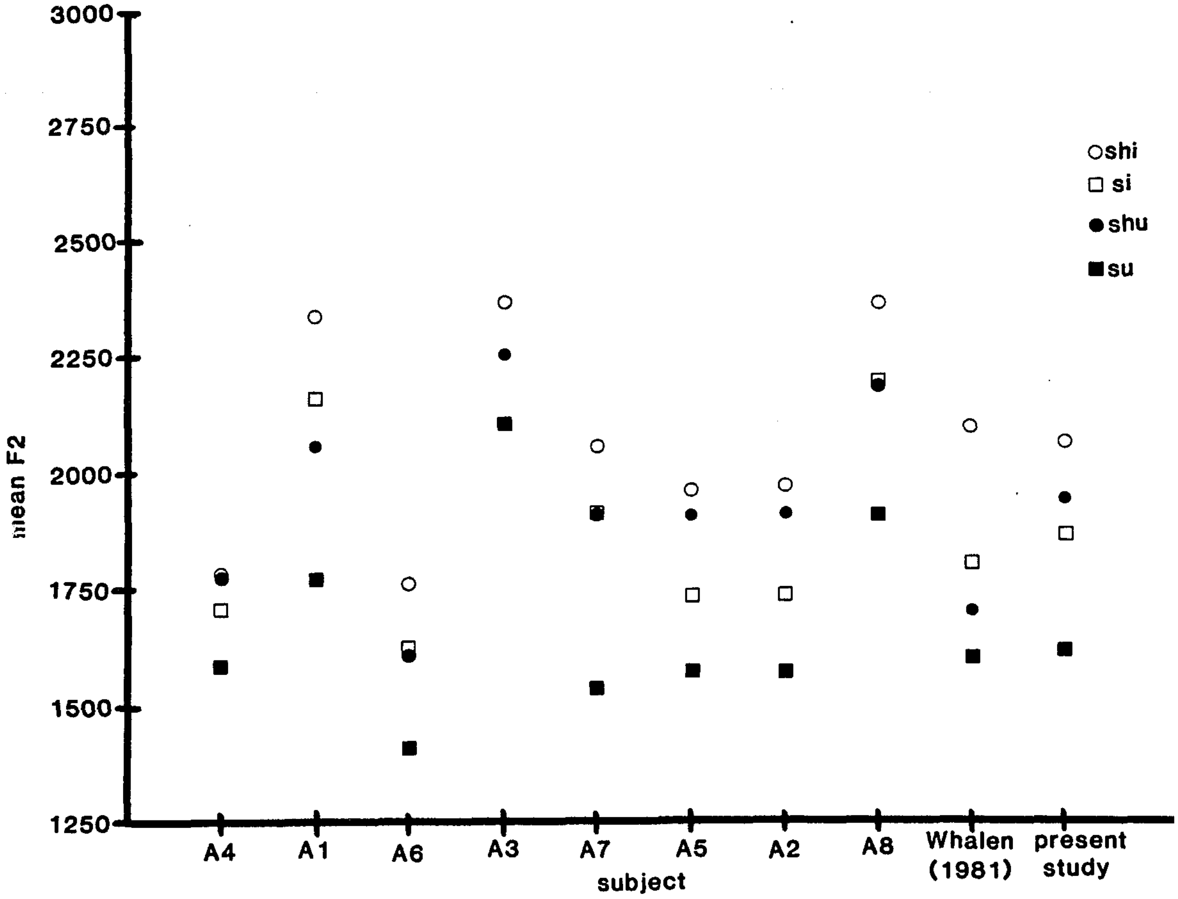
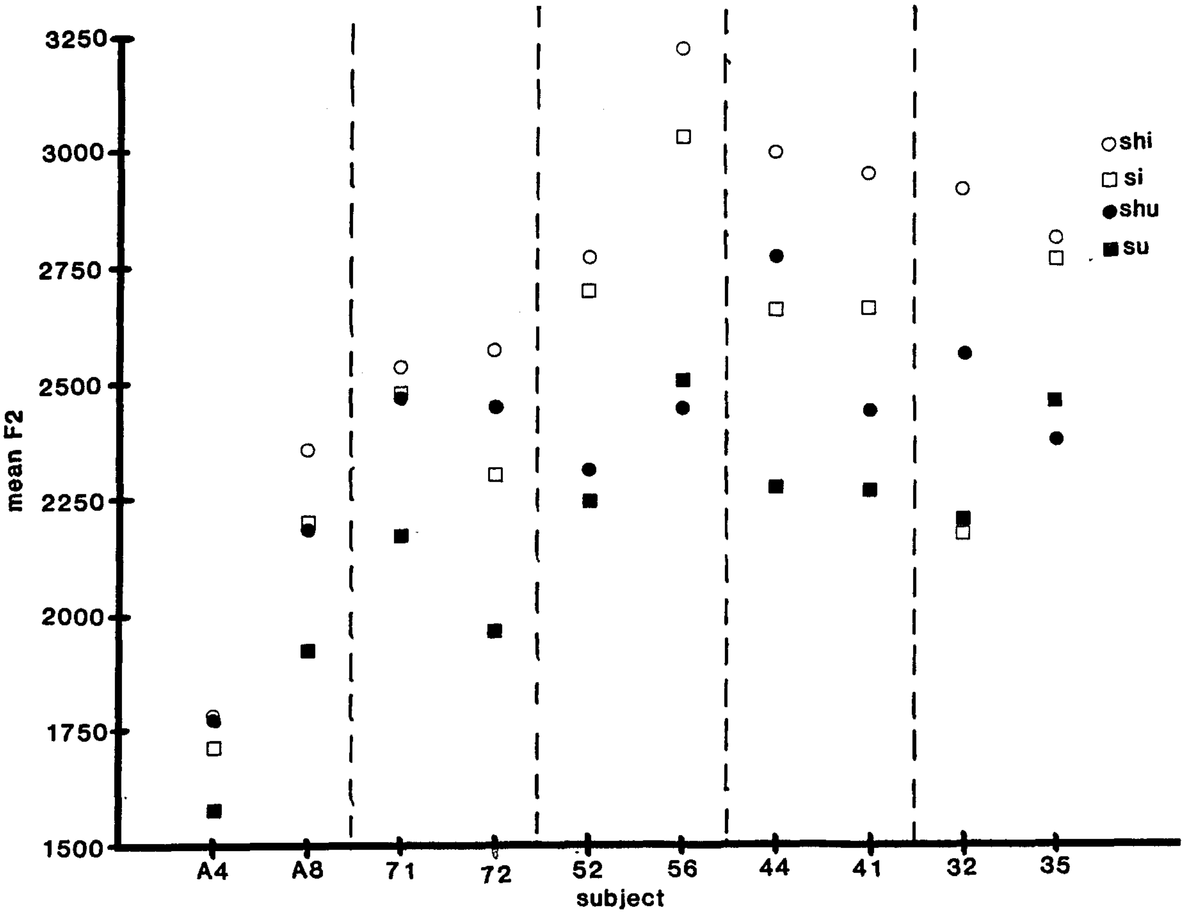


Figure 28: Mean F2 values for selected subjects at VO2



DISCUSSION

The adults participating in this study demonstrated one effect that could not have been predicted from previous studies (e.g., Heinz & Stevens, 1961; Kunisaki & Fujisaki, 1977): mean centroids for /ʃ(u)/ were higher than for /ʃ(i)/. What this may indicate is that the constriction was not sufficiently narrow in adults' productions of /ʃ/ to prevent coupling of the back cavity to the source. Because the tongue is further back for /u/ than for /i/, back cavity resonances are higher for /u/ than for /i/. These back cavity resonances appear to influence the acoustic results of /ʃ/ production for adults. However, differences in centroids for children's productions of /ʃ(i)/ and /ʃ(u)/ were in the direction predicted by earlier work: /ʃ(i)/ centroids were higher than /ʃ(u)/ centroids. It seems that the constriction was sufficiently narrow in children's /ʃ/ productions to prevent the coupling of the back cavity to the source.

For all subjects in this experiment, the distinction between /s/ and /ʃ/ was greater in the /i/ context than in the /u/ context. One possible explanation for this involves the generally expected results of liprounding. In the /u/ context, both /s/ and /ʃ/ are produced with liprounding; in the /i/ context, only /ʃ/ is produced with liprounding. Therefore, greater differences between fricatives produced in the /i/ context would be predicted.

In general, children in this experiment demonstrated smaller differences between /ʃ/ and /s/. This fact is evident both in the fricative spectra which show less distinct patterns for /ʃ/ and /s/ tokens produced by children than for those produced by adults, and in the centroid values which differ less between /ʃ/ and /s/ for children than for adults. In addition, the children showed more evidence of vowel identity within the fricative portion of the syllable. These two findings (less distinct fricatives and greater vowel effects) are both reflected in the generally flatter fricative spectra obtained for children as compared to adults (Appendix E). The question is why would these two results be obtained for children's productions?

Menn (1978) argues that young children are sensitive to the elements in a syllable before they are able to combine them with the correct temporal relations. If this is so, and if we further assume that motor control would be poorer for young children than for adults, we may conclude that the children in this study "knew" that they needed to produce fricative plus vowel, but were not able to do so with the temporal precision of adults. This lack of precision would lead to both the less distinct fricatives and the greater vowel effects found.

Of course, children did not show a total lack of attention to temporal organization since their productions obviously consisted of a predominantly fricative segment followed by a predominantly vowel segment. In fact, in some samples children demonstrated evidence of temporally separating the syllabic components more than adults did. This temporal segregation of segments is illustrated in the waveforms.

Figure 29 shows a sample waveform for the region of an adult's utterance between VO2-60 ms and VO2+150 ms. From this waveform, it appears as if the frication continues to vocalic onset, but the two segments remain discrete. This production pattern requires good motor control to achieve the precise temporal relations between laryngeal and supralaryngeal structures. Figure 30 is a sample waveform of the same region taken from a child's utterance. This pattern was common in children's speech and shows that, at least some of the time, they separated temporally fricative and vowel to even a greater extent than adults did. The amplitude of the frication in this sample decreases much sooner before vocalic onset than in the adult's sample. Figure 31 shows another, somewhat less common, waveform obtained from the same child producing the same utterance. This waveform shows that the child's attempt to cease frication exactly at vocalic onset was not successful: here, frication evidently continues during the vocalic portion. However, Figure 32, another token of the same utterance from the same child, illustrates that sometimes these attempts were successful. In this waveform, frication terminates precisely at vocalic onset, as in adults' productions. In general, then, the temporal structure of children's productions was more variable than adults'. One possible explanation for this increased variability could be that children lack the motor control required to open the constriction sufficiently to end frication precisely as voicing begins. However, the pattern of increased temporal separation shown for the child's sample in Figure 30 suggests that, at least sometimes, attempts were made to separate the segments in spite of this lack of motor control.

Given this increased variability in the temporal structure of children's productions over adults', we might also expect more variability in the spectral structure of children's than adults' productions. Accordingly, standard deviations (S.D.) of centroids for samples taken at

Figure 29: Fricative-to-vowel acoustic segment for adult

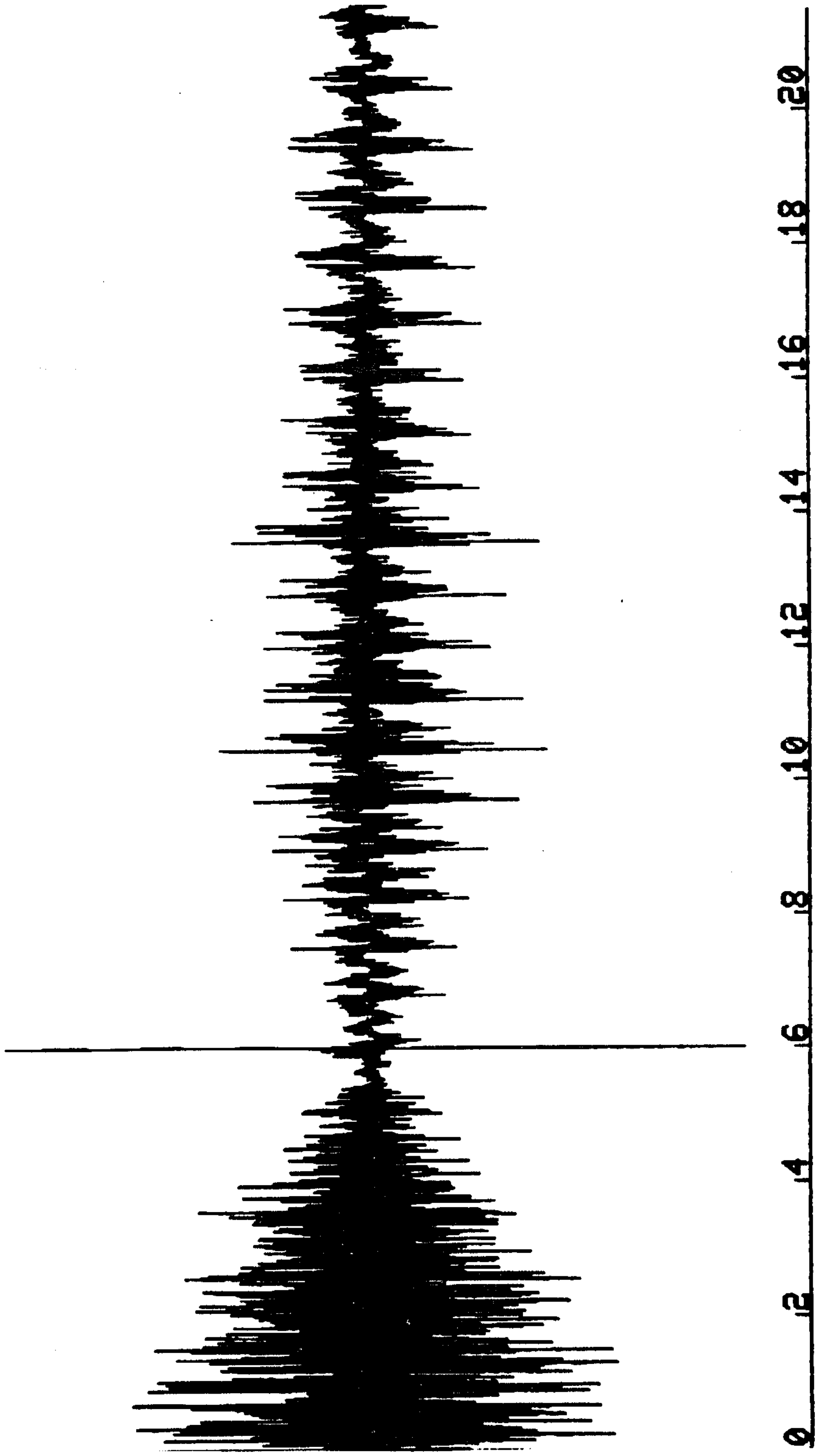


Figure 30: Fricative-to-vowel acoustic segment for child, temporally separated

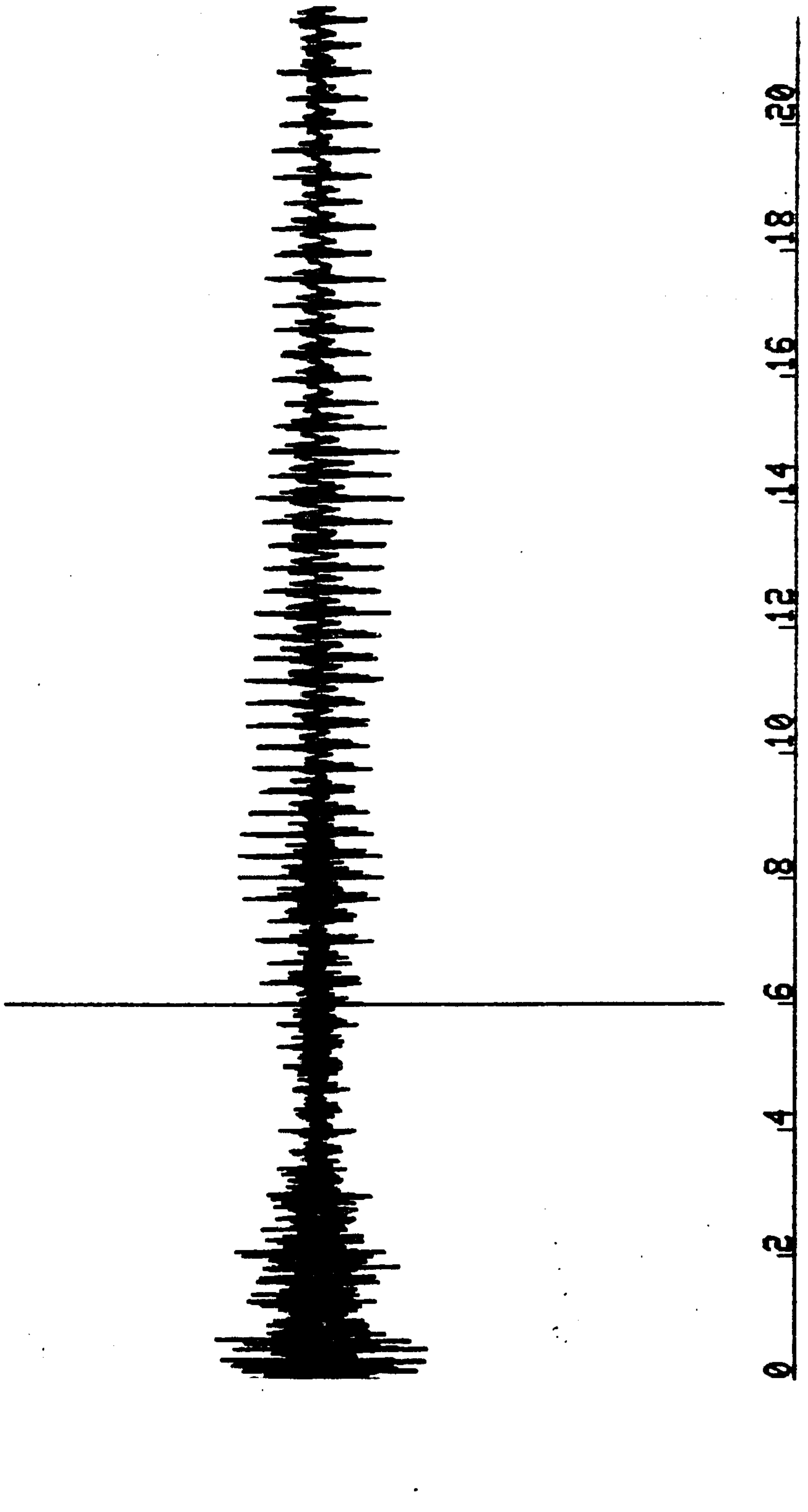


Figure 31: Fricative-to-vowel acoustic segment for child, temporally overlapping

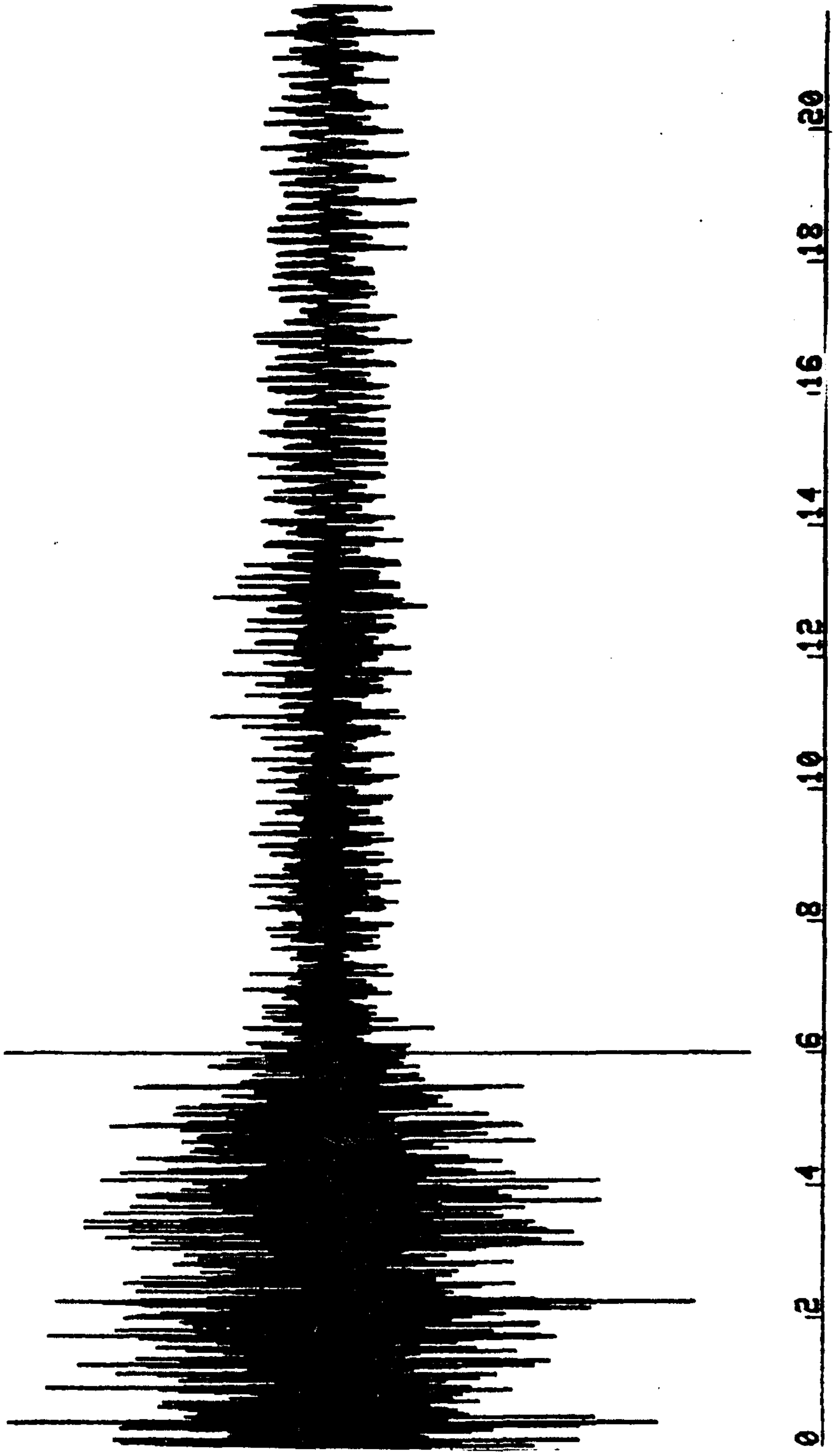
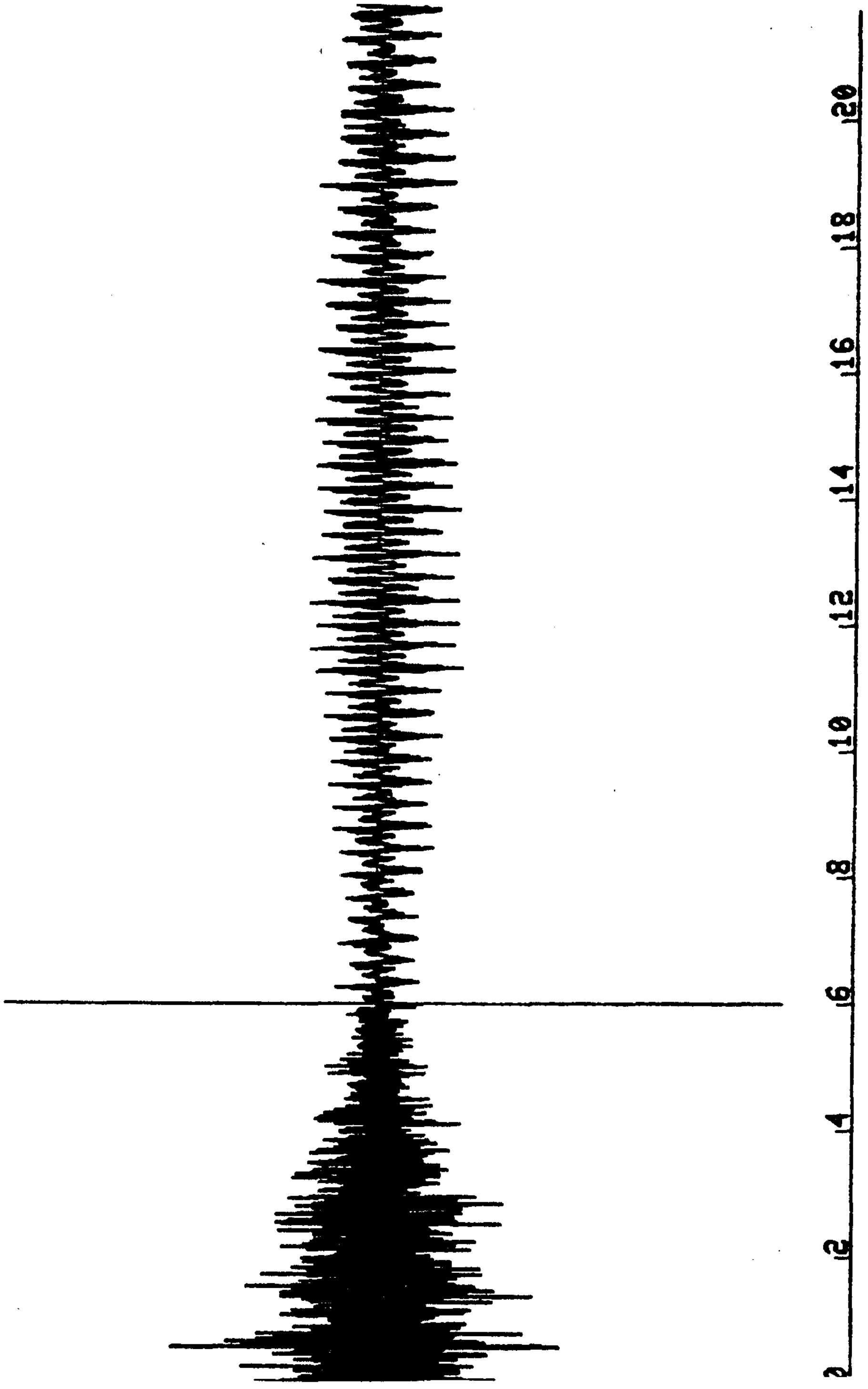


Figure 32: Fricative-to-vowel acoustic segment for child, temporally appropriate



VO1-100 ms were computed, and are listed in Table 26. The percentage values indicate the percentage of mean centroid values (taken from Table 8) represented by the mean S.D. for adults and children. For example, the mean adult S.D. of 252 Hz for /ʃ(i)/ represents 4% of the mean adult centroid value of 5653 Hz. It can be seen from these percentages that children were slightly more variable than adults, but the differences were not great. Thus, while inspection of waveforms indicates that the children were more variable than the adults in the temporal domain, measurements of spectral structure demonstrated scarcely more variability for the children than for the adults in the spatial domain.

This combination of results suggests that the children were not simply trying to produce the same articulatory targets as adults, and the spatial targets were altered in the children's productions by output constraints. If results were primarily due to output constraints, more spectral variability would probably have been demonstrated. Rather, it seems that these children actually had different articulatory targets than adults, and that the targets were altered by output constraints

Table 26: Standard deviations of centroids for samples taken starting at VO1-100 ms

Adults	ʃ(i)	ʃ(u)	s(i)	s(u)
A4(most)	325	308	336	379
*A1(2)	222	334	180	310
A6(3)	354	199	403	505
*A3(4)	242	276	165	319
A7(5)	301	277	311	254
A5(6)	162	258	230	277
*A2(7)	266	314	268	178
*A8(least)	141	146	269	217
$\bar{X}$=	252	264	270	305
	(4%)	(5%)	(4%)	(4%)
Children				
71(most)	326	402	405	373
72(least)	273	433	272	322
52(most)	257	525	408	220
56(least)	520	275	381	374
44(most)	305	340	231	270
41(least)	581	528	565	753
32(most)	345	407	417	488
35(least)	240	196	169	411
$\bar{X}$=	356	388	356	401
	(6%)	(6%)	(5%)	(6%)

far less than might have been expected.

The picture of the children's production that emerges from the results reported here is one in which the child is, on the one hand, coproducing fricative and vowel, and on the other, attempting to segregate them temporally. In other words, while young children 3 to 7 years old may still be speaking "syllabically" somewhat as they did in their earliest attempts at words, they have become sensitive to the segmental structure of the syllable and are attempting to instantiate it in their own speech. This account is compatible with Menn's (1978) contention that children are sensitive to the different elements in a syllable (e.g., frication plus vowel) and their approximate positions, but cannot quite combine them appropriately.

This account is also compatible with the results of the perception experiment. The overall temporal structure of syllables produced by these children was approximately correct: fricative always preceded vowel. But the finer temporal characteristics were variable across the children's productions. In the perception experiment, children did not seem to use the information provided by the fricative noise spectrum as reliably as the adults. Selective attention to a specific portion of the syllable in perception may call on capacities related to those needed to terminate frication precisely at vowel onset in production, a skill which these children apparently lacked.

The finding that children demonstrate a greater vowel effect in fricative production than adults conflicts with Kent's (1983) claim that young children coarticulate less. This claim was based on the observation that vocalic transitions were less extensive in certain children's productions than in adults', as illustrated also here in the sample spectrograms (Appendix F). However, the reason children's vocalic transitions are less extensive than adults' would seem to be, not that they coarticulate less, but that they coarticulate (or coproduce) more. Children's tongues are evidently closer to the vowel target during consonant production, so that the tongue does not have as far to move once the vocalic portion is initiated. This coproduction evidently coexists with a counter tendency to segregate the segments --- in which they are not always successful because they lack the motor skill needed to make fluent switches from a constriction to an open vocal tract.

Actually, Kent (1983), in the same paper in which he argues that children coarticulate less than adults, also proposes a principle of "everything moves at once" (p. 70) to capture the idea that individuals who have difficulty with the motor control required for normal production (e.g., young children and the motorically impaired) may start producing all the syllabic elements together at the beginning of each syllable. However, Kent does not offer this production pattern as evidence for greater coarticulatory effects in the speech of children.

The 7-year-olds in this study demonstrated essentially the same patterns of production as the younger children. However, in the perception experiment, 7-year-olds performed similarly to adults. From this we may infer that the ability to recover individual segments in perception precedes the ability to produce these segments with motor and temporal precision.

Comparisons for individual results of coarticulatory effects demonstrated in perception and production show no correlation in the sizes of these effects. There was some evidence of a relation between mean values: The mean sizes of vowel effects found for adults in perception were similar to the mean sizes of vowel effects demonstrated in adults' productions. However, no firm conclusions can be drawn from this because of the variability in the magnitude of the vowel effects demonstrated in adults' productions.

Finally, the results reported here demonstrated that vowel effects were evident at least 100 ms before vocalic onset for all subjects. There was no region of static, invariant acoustic information in these fricative segments. This result is particularly interesting because fricatives were described by a measure encompassing the entire spectrum: the centroid.

GENERAL DISCUSSION

The role of coarticulation in the production and perception of speech by young children was investigated to try to extend our understanding of the role of coarticulation in adult speech. Evidence of coarticulation was found in children's production, and sensitivity to its acoustic consequences was demonstrated in their perception. In fact, in certain respects these effects were stronger in children's perception and production than in adults'.

On the face of it, though, the observed coarticulatory effects in children's perception and production seem paradoxical. In the perception experiment, the children appeared to weight the vowel context less heavily than adults, and the transitional information more heavily. On the other hand, the production results indicate that the children exhibited larger vowel effects than adults, but reduced transitions. This paradox may be resolved by considering the source of each effect.

The source of the vowel effect is different for perception and production of speech. In perception, the presumed source of the vowel effect is the listener's tacit knowledge of the acoustic consequences on the fricative of anticipatory liprounding for the vowel. A prerequisite for development of this tacit knowledge must be that the listener's attention somehow be focused on the temporally distinct segments of the syllable. In production, though, an effect of the vowel on the fricative suggests that the speaker is attempting to produce the syllable as a single unit rather than as a pattern composed of individual segments. Thus, if we assume that 3- and 4-year-old children do not selectively attend to individual segments in perception or production, and that even 7-year-olds still do not do so in production, as much as adults do, we have a unified account of the perception and production vowel effects that resolves the paradox.

Moreover, we can comfortably extend this account to explain the observed effects on the transition. The children's heavy weighting of transitional information in perception reflects sensitivity to syllabic structure, and the reduced transitions in production reflect exactly the same sensitivity: Transitions are reduced precisely because vowel and consonant are coproduced to a greater extent, so that there is less formant movement once voicing begins.

While the relative sizes of the vowel and transition effects for children compared to adults both suggest that children do not perceive or produce speech in as segmental a manner as adults, it is clear that these 3- to 7-year-olds were not using the syllabic approach to speech of younger children. For example, these children were able to use the fricative spectrum to a considerable extent in perception; otherwise, identification functions would have been much flatter. Also, they appear to have frequently attempted a temporal separation of segments in production. Thus, both perception and production results suggest that these children have moved away from the whole unit approach used by younger children, yet still do not perceive or produce speech with as much attention to individual segments as adults. Furthermore, the different results demonstrated by 7-year-olds for the perception and production experiments suggest that an individual segment approach is attained for perception before production. The requirements of production and perception evidently constrain the child's development in different ways at different times. The eventual result of development for both, however, is a system delicately balancing the need for accurate production and perception of individual segments, with the need for efficient combination of these segments into coarticulated syllables.

Implications for theories of coarticulation

The perception results reported here clearly argue against the acoustic theory of speech perception proposed by Stevens and Blumstein (1978). If listeners used innate property detectors to perceive speech based on static, invariant characteristics of the signal, an opposite pattern of results would be expected. That is, children, more than adults, would display evidence of identifying consonants based on the "primary" consonantal portion of the signal, and of disregarding the "secondary" transitional information. Instead, the results showed that all subjects, including the youngest children, drew on all available information to make their phonetic decisions. We can therefore dismiss the disposing account of speech perception.

The present results also fail to support the speech code view of coarticulation. This theory assumes that the speaker/listener has tacit knowledge of the underlying phonemic categories. The speaker is said to execute a string of phonemic segments, restructuring them at output for the sake of efficiency, while the listener uses his/her tacit knowledge of how the segments were restructured to recover the individual segments. For this theory to have been supported, evidence would have had to be found that children are more sensitive to the underlying phonemic structure than they were. For example, if it had been found that children demonstrated the same sensitivities to coarticulatory effects in perception as adults, but weaker coarticulatory effects in production, it might have been argued that children recognize the underlying phonemic structure of speech as well as adults, but that their production is altered by output constraints. However, the perceptual results suggest that children are not as sensitive to the phonemic structure as adults.

The dynamic theory of coarticulation (or coproduction) receives the strongest support from these results. The findings reported here could indicate that children between the ages of 3 and 7 are beginning to separate those aspects of the signal which are the vowel and consonant in perception, and are trying to incorporate what they know of these separate entities into production. However, there is one contradiction between these results and the predictions of the dynamic theory of coarticulation.

According to this theory, the adult listener actually perceives the segments separately. The acoustic results of producing each segment are combined in the signal, but the patterns that represent each segment remain perceptually separate, because they are qualitatively distinct. In this study, though, the spectral characteristics of the fricative noises were not influenced by vowel coproduction because synthetic stimuli were used. Thus, there was no vowel information to be separated from the fricative. Yet, vowel context effects were demonstrated. These effects seem to require that the listener somehow "infer" fricative identity from characteristics of the vocalic segment --- perhaps from tacit knowledge of the effects of vowel liprounding, as Fujisaki and Kunisaki (1977) originally proposed. However, the idea of making inferences from one segment to another contradicts the concept of directly perceiving the separate aspects of the signal as individual segments, as suggested by the dynamic theory of coarticulation. Perhaps it is the case, then, that a listener approaching a speech perception task using synthetic stimuli automatically attempts the sort of analysis described by Fowler et al. (1980), separating the synthetic signal into components which as closely as possible resemble those which might be found in natural speech.

Implications for future research

The acoustic analyses described here offer the clearest suggestions for future research. Children's speech samples demonstrated a spectral pattern indicating extensive coproduction of segments, but a temporal pattern indicating difficulty in combining these segments appropriately. Several kinds of studies could be conducted to help describe these patterns in more detail. First, a perceptual study in which (adult) listeners were asked to identify vowel quality from fricatives edited from children's and adults' productions of fricative-vowel syllables would help determine whether or not increased vowel effects found in the acoustic analysis of children's fricatives are perceptually salient. Also, an investigation of children's speech timing using a more direct measure than acoustic analysis (e.g., comparison of measurements made from light emitting diodes attached to different articulators) would provide a more-detailed description of timing in children's speech.

Summary

Four possible outcomes for this study were described in the introduction. Possibility #4 was that young children would demonstrate coarticulatory effects in production, and sensitivity to these effects in perception. This was the finding obtained. Children as young as 3 years showed evidence of coarticulation, and attended to its acoustic consequences. Thus, coarticulation seems to be an essential quality of speech. However, two nuances were found that would not have been predicted on the basis of previous work. First, differences in magnitude for different coarticulatory effects were demonstrated in children's production, and for children's sensitivities to these effects in perception. Second, in both production and perception, children demonstrated greater effects of one kind or another than did adults. Both of these nuances may be accounted for by a description of speech development in which the child originally attends to the syllabic structure as a whole and only gradually discovers the phonemic units represented in the syllable.

Notes

[1] The software program used to do these probit analyses would not run if less than 2 stimulus levels had proportion scores between .10 and .90. Smaller step sizes, within the boundary region would have been needed to obtain enough data to fit a curve. Therefore, slope values could not be obtained in these cases. This presented a problem for analysis because elimination of these points would have artificially deflated mean slope values. To substitute average values based on the scores of other subjects in each respective age group would have similarly decreased slopes. The most appropriate choice seemed to be to substitute the highest slope obtained for this set of stimuli, which was 2.16. This value was obtained when proportion scores for 2 points were between .10 and .90. Slope values substituted in this way are indicated in Appendix C by an asterisk. While this method deflated mean slope values for the group somewhat, the effect was not as great as it would have been with any other method.

The probit analysis program used did a goodness-of-fit test to estimate how well a cumulative normal distribution fit the data for each context. Consequently, chi-square values were obtained for each subject for each context. Information about the probability of obtaining each chi-square value under the null hypothesis that a cumulative normal distribution fit the data are provided in Appendix D. Probabilities of less than .01 are represented with an asterisk; exact probabilities between .01 and .05 are provided.

[2] For centroids, it was not clear whether overall means or separate means for males and females should be provided because a complete description of the factors contributing to fricative spectra does not exist. All 3 values are given for comparison in the tables of mean centroids (Tables 8 and 9), but tables of difference scores (Tables 10 and 11) give only overall means. For tables listing F2 values and F2 difference scores, only separate means for males and females are given because it was clear *a priori* that differences would be expected due to differences in vocal tract length, and the strong relation between formant values and vocal tract length.

All tables listing ratios and Eta^2 values provide only overall means because any individual differences due to vocal tract length would have been taken into account.

Only overall means for children are provided in all tables because (1) no systematic variation is apparent, and (2) there did not seem any reason to do so.

Appendix A: Individual phoneme boundaries

	/(ʃ)i/	/(s)i/	/(ʃ)u/	/(s)u/
Adults				
A1	4.60	4.40	5.34	2.65
A2	5.66	5.03	6.00	4.74
A3	6.92	6.50	6.46	4.53
A4	6.16	5.41	9.62	2.45
A5	4.65	4.40	5.50	3.79
A6	4.82	4.25	5.36	2.96
A7	4.32	3.80	5.00	2.98
A8	6.50	6.40	6.00	6.00
7-year-olds				
71	4.39	3.24	5.50	1.65
72	4.76	4.40	5.50	4.40
73	5.00	5.00	6.00	4.03
74	5.70	5.40	5.91	4.09
75	5.03	4.60	5.50	3.60
76	6.33	4.52	6.22	4.47
77	5.49	5.21	5.40	3.71
78	4.98	4.70	6.44	4.00

Appendix A (continued)

	/(ʃ)i/	/(s)i/	/(ʃ)u/	/(s)u/
5-year-olds				
51	6.26	5.32	7.87	4.96
52	4.50	4.09	5.76	1.86
53	5.79	5.30	7.40	5.79
54	6.00	4.88	5.88	4.17
55	5.81	4.31	5.93	3.25
56	5.46	4.73	5.98	4.89
57	5.03	5.38	4.75	6.50
58	5.38	4.75	6.50	3.12
4-year-olds				
41	5.70	4.97	6.41	4.22
42	6.46	5.87	10.89	5.57
43	5.05	4.00	5.76	3.49
44	6.00	5.32	9.42	4.52
45	6.70	6.29	7.29	4.84
46	5.65	4.41	6.73	4.37
47	6.42	5.50	8.47	4.57
48	5.58	5.41	8.17	4.53

Appendix A (continued)

	/(ʃ)i/	/(s)i/	/(ʃ)u/	/(s)u/
3-year-olds				
31	5.02	4.27	4.89	0.11
32	4.31	4.27	6.68	0.64
33	4.46	5.14	6.74	3.16
34	4.96	4.13	7.84	4.47
35	6.50	5.25	6.33	5.23
36	6.44	4.16	8.71	5.93
37	4.85	4.18	10.91	5.81
38	5.09	5.36	10.82	5.93

Appendix B: Individual slopes

	/(ʃ)i/	/(s)i/	/(ʃ)u/	/(s)u/
Adults				
A1	2.16	2.16	1.07	0.97
A2	1.21	1.67	2.16*[a]	1.40
A3	0.82	1.81	0.99	0.93
A4	0.81	1.00	0.29	0.77
A5	1.45	0.86	1.46	1.79
A6	1.37	1.04	0.96	0.85
A7	1.09	0.84	2.16*	1.17
A8	1.81	2.16	2.16*	2.16*
7-year-olds				
71	0.52	0.77	1.81	0.95
72	1.69	2.16	2.16	2.16
73	2.16*	2.16*	2.16*	1.67
74	1.93	0.74	1.05	1.35
75	1.67	1.62	1.81	2.16
76	0.93	0.70	1.43	1.07
77	0.72	1.79	2.16	1.10
78	0.80	1.93	1.24	2.16*

[a] Asterisk indicates that this value was substituted because it was not possible to do probit analysis; actual value would be greater.

Appendix B (continued)

	/(ʃ)i/	/(s)i/	/(ʃ)u/	/(s)u/
5-year-olds				
51	0.92	0.60	0.91	0.73
52	1.81	0.60	0.93	0.69
53	1.79	1.93	0.78	1.79
54	2.16*	1.71	1.71	1.37
55	1.20	1.51	0.86	1.00
56	0.94	0.55	0.84	0.89
57	1.67	0.63	1.17	0.99
58	0.71	1.19	0.84	1.18
4-year-olds				
41	0.80	1.67	0.17	0.32
42	1.32	0.78	0.19	0.36
43	0.52	2.16*	0.38	0.70
44	2.16*	0.81	0.37	0.64
45	0.31	0.23	0.26	0.28
46	0.57	0.82	0.37	0.55
47	0.49	0.90	0.98	0.86
48	0.72	0.64	0.64	1.07

Appendix B (continued)

	/(ʃ)i/	/(s)i/	/(ʃ)u/	/(s)u/
3-year-olds				
31	0.27	0.29	0.29	0.25
32	0.50	0.49	0.24	0.15
33	0.42	0.45	0.35	0.37
34	0.47	0.60	0.63	0.72
35	0.75	0.65	0.52	0.47
36	0.37	0.33	0.37	0.89
37	0.73	0.51	0.31	0.64
38	0.28	0.66	0.21	0.65

Appendix C: Probabilities of chi-square values obtained for tests of goodness-of-fit of data to cumulative normal distribution

	(ʃ)i	(s)i	(ʃ)u	(s)u
Adults				
91				
92			NA[a]	
93				
94				
95		*[b]		
96				
97	*	.02[c]	NA	
98			NA	NA
7-year-olds				
71	.01			
72				
73	NA	NA	NA	
74		*		
75				
76		.04		
77	*			
78	*			NA

[a] Not applicable because subject did not have 2 or more points on distribution between .10 and .90 in this context, therefore program would not perform probit analysis

[b] Indicates that $p<.01$

[c] Exact p values given when between .01 and .05

Appendix C (continued)

	(ʃ)i	(s)i	(ʃ)u	(s)u
5-year-olds				
51	*	*		
52		*		
53				
54	NA			
55				
56		*	.01	*
57		*		
58	.02			.01
4-year-olds				
41	.03			
42				
43		NA		
44	NA	*		
45	.01			
46	*	*	*	.01
47	*			
48	.02	.03		

Appendix C (continued)

	(ʃ)i	(s)i	(ʃ)u	(s)u
3-year-olds				
31				
32	.03			
33				
34	.02	*		*
35		*	.01	
36			*	
37		.01		
38				*

Appendix D: Individual difference scores

	Transition effects		Vowel Effects	
	(ʃ)i-(s)i	(ʃ)u-(s)u	(ʃ)i-(ʃ)u	(s)i-(s)u
Adults				
A1	0.20	2.69	-0.74	1.75
A2	0.63	1.26	-0.34	0.29
A3	0.42	1.93	0.46	1.97
A4	0.75	7.17	-3.46	2.96
A5	0.25	1.71	-0.85	0.61
A6	0.57	2.40	-0.54	1.29
A7	0.52	2.02	-0.68	0.82
A8	0.10	0.00	0.50	0.40
7-year-olds				
71	1.15	3.85	-1.11	1.59
72	0.36	1.10	-0.74	0.00
73	0.00	1.97	-1.00	0.97
74	0.30	1.82	-0.21	1.31
75	0.43	1.90	-0.47	1.00
76	1.81	1.75	0.11	0.05
77	0.28	1.69	0.09	1.50
78	0.28	2.44	-1.46	0.70

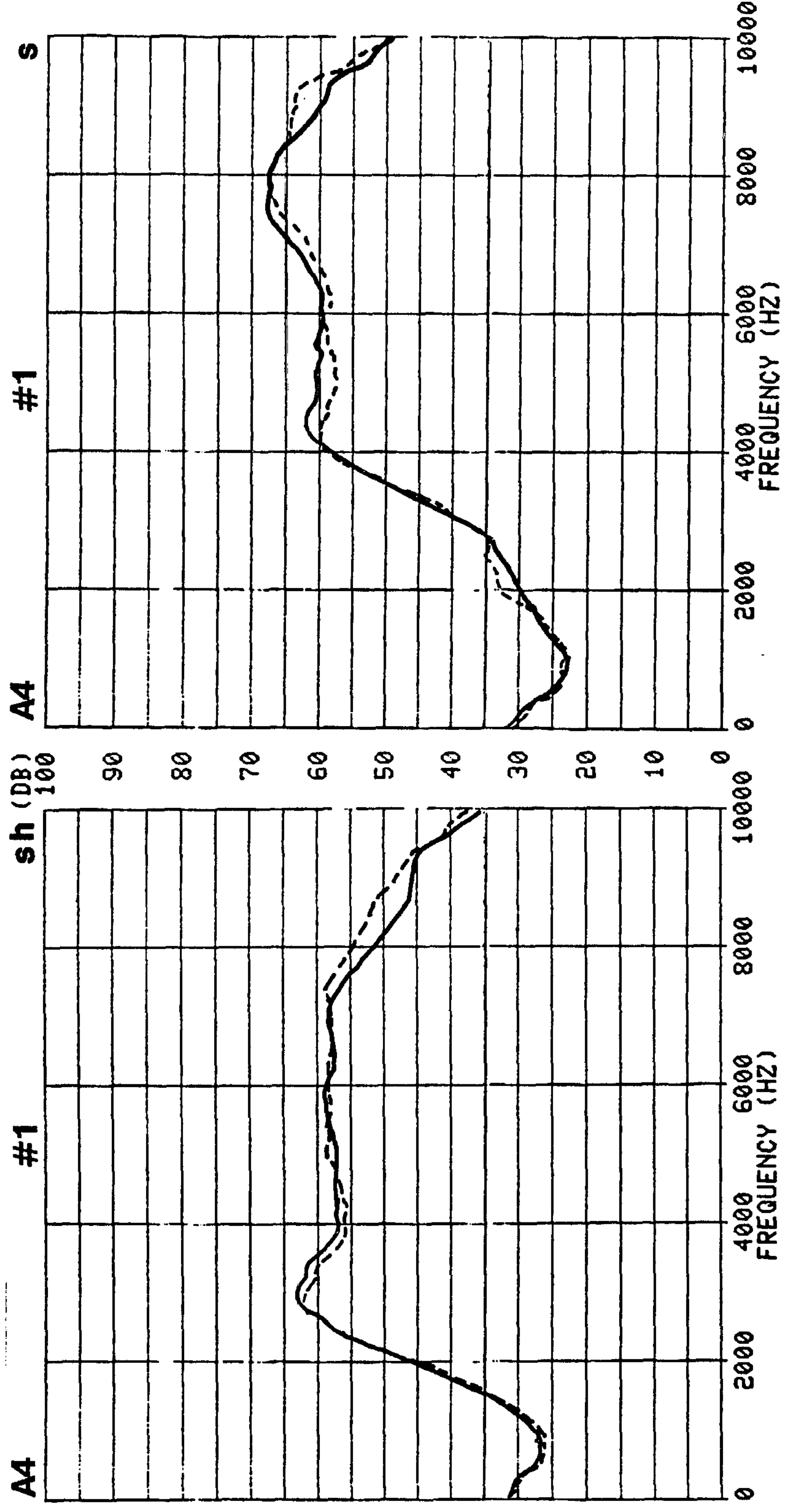
Appendix D (continued)

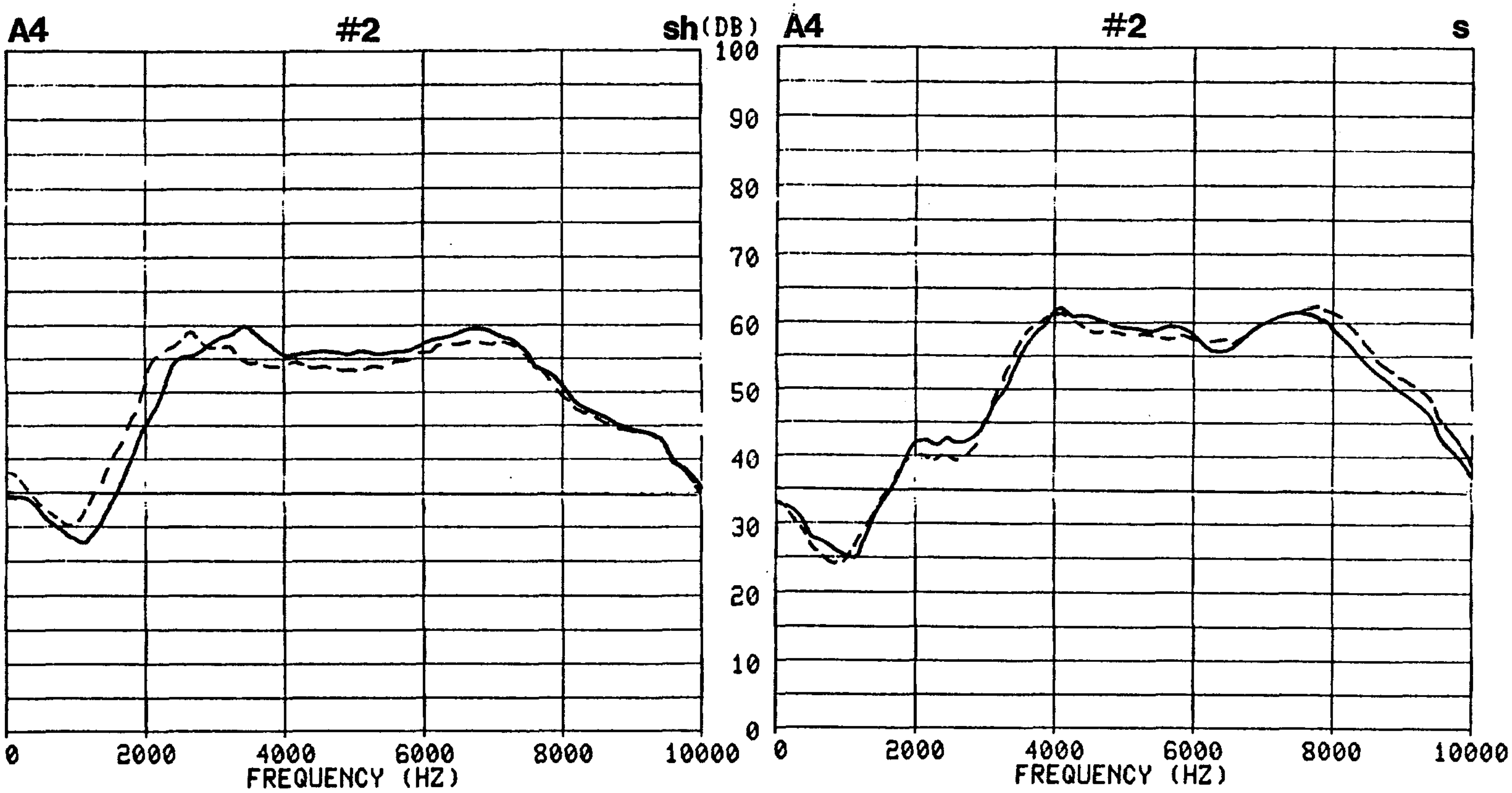
	Transition Effects		Vowel Effects	
	(ʃ)i-(s)i	(ʃ)u-(s)u	(ʃ)i-(ʃ)u	(s)i-(s)u
5-year-olds				
51	0.94	2.91	-1.61	0.36
52	0.41	3.90	-1.26	2.23
53	0.49	1.61	-1.61	-0.49
54	0.67	1.71	-0.33	0.71
55	1.50	2.68	-0.12	1.06
56	0.73	1.09	-0.52	-0.16
57	0.75	2.11	-1.81	-0.45
58	0.63	3.38	-1.12	1.63
4-year-olds				
41	0.73	2.19	-0.71	0.75
42	0.59	5.32	-4.43	0.30
43	1.05	2.27	-0.71	0.51
44	0.68	4.90	-3.42	0.80
45	0.41	2.45	-0.59	1.45
46	1.24	2.36	-1.08	0.04
47	0.92	3.90	-2.05	0.93
48	0.17	3.64	-2.59	0.88

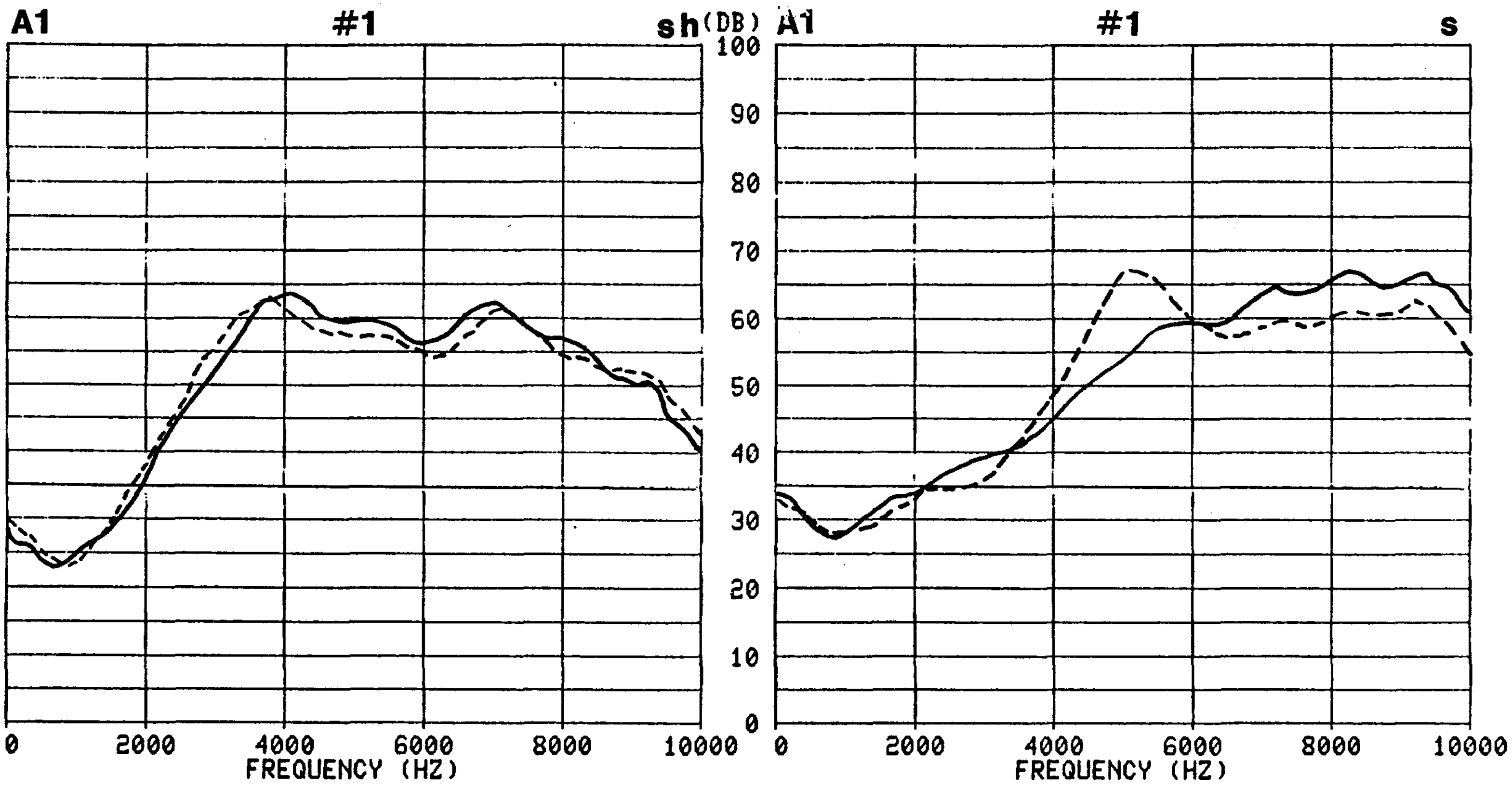
Appendix D (continued)

	Transitions Effects		Vowel Effects	
	(ʃ)i-(s)i	(ʃ)u-(s)u	(ʃ)i-(ʃ)u	(s)i-(s)u
3-year-olds				
31	0.75	4.78	0.13	4.16
32	0.04	6.04	-2.37	3.63
33	-0.68	3.58	-2.28	1.98
34	0.83	3.37	-2.88	-0.34
35	1.25	1.10	0.17	0.02
36	2.28	2.78	-2.27	-1.77
37	0.67	5.10	-6.06	-1.63
38	-0.27	4.89	-5.73	-0.57

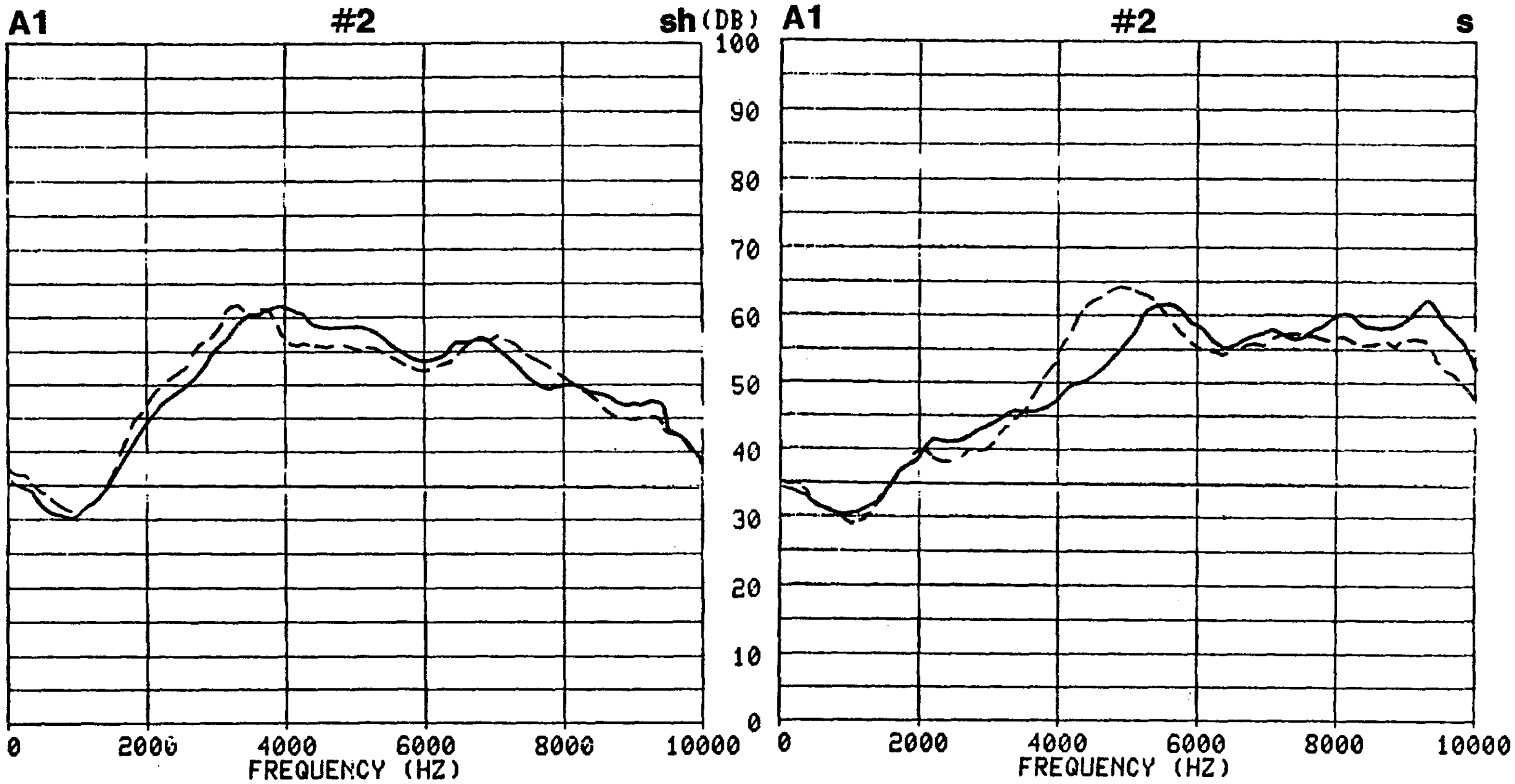
Appendix E: Averaged spectra

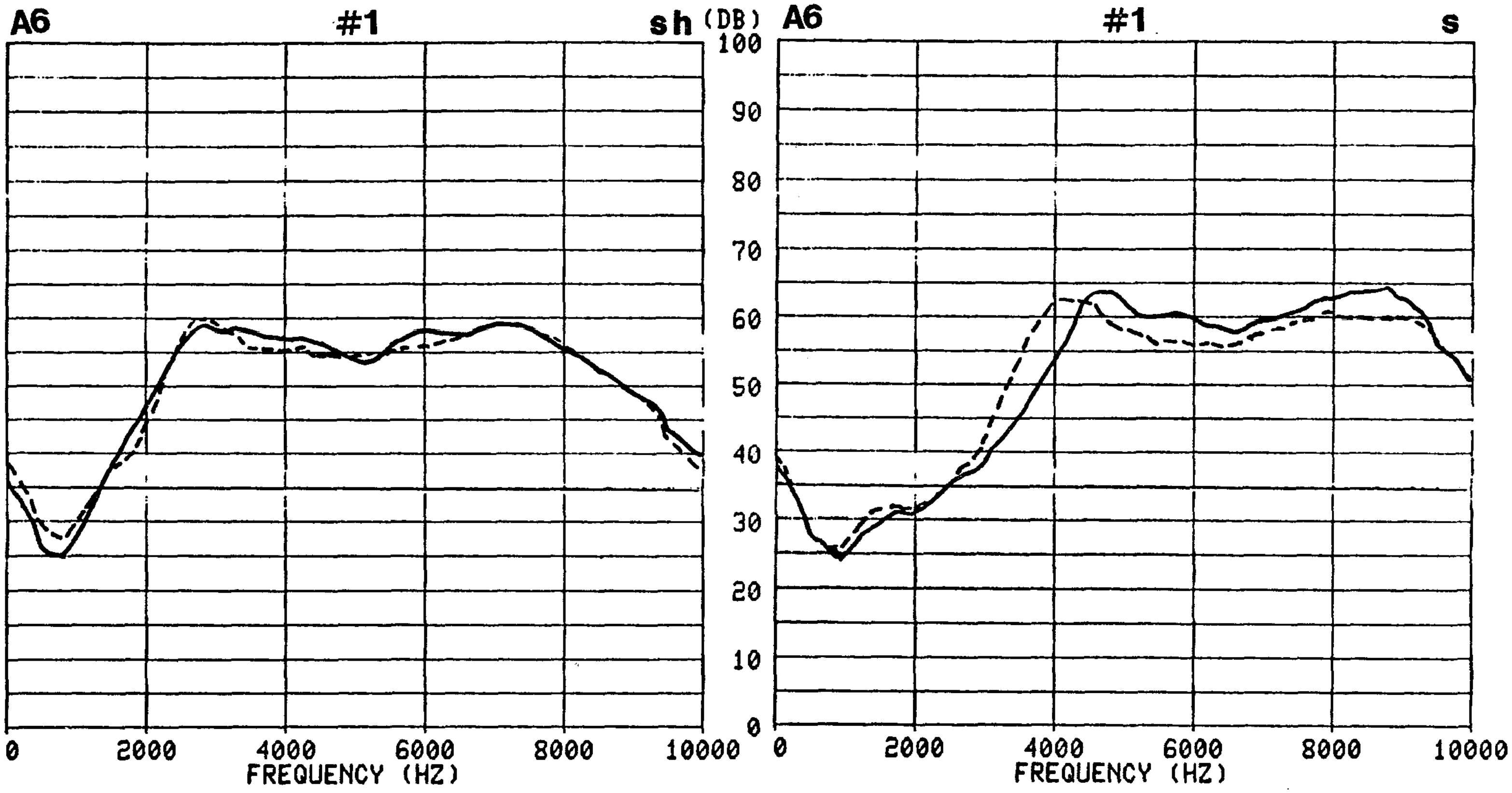


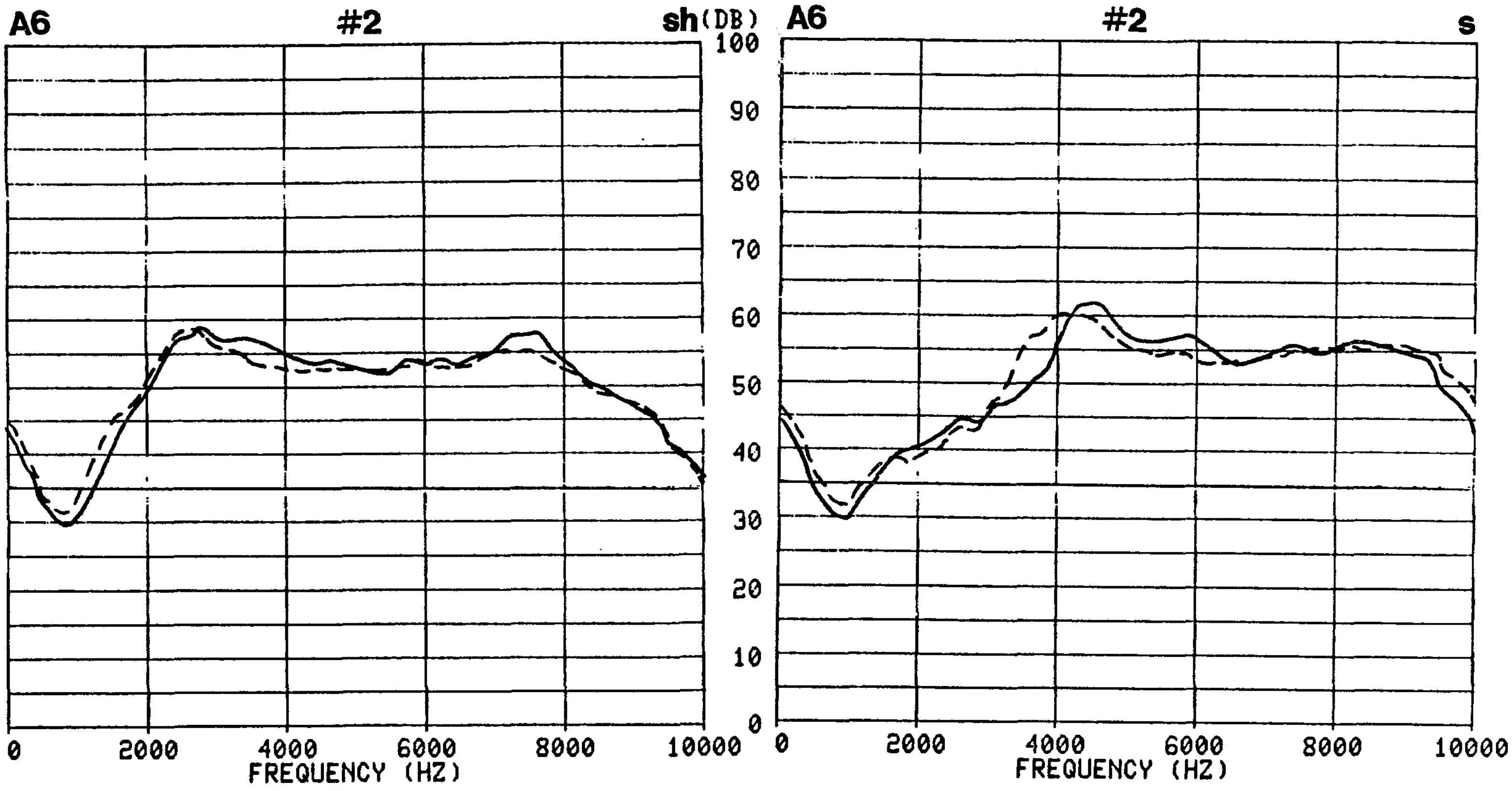

A4
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
A4
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

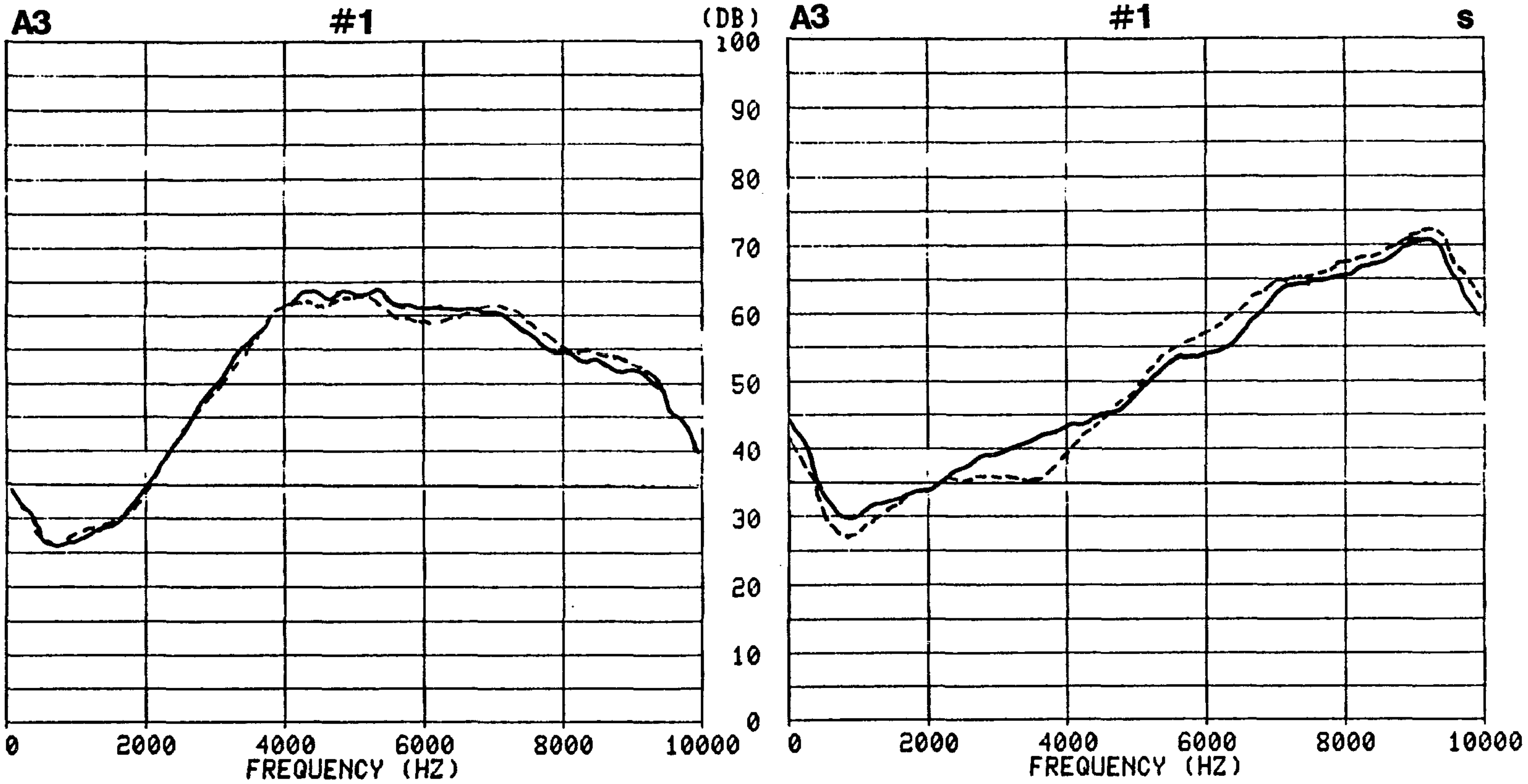


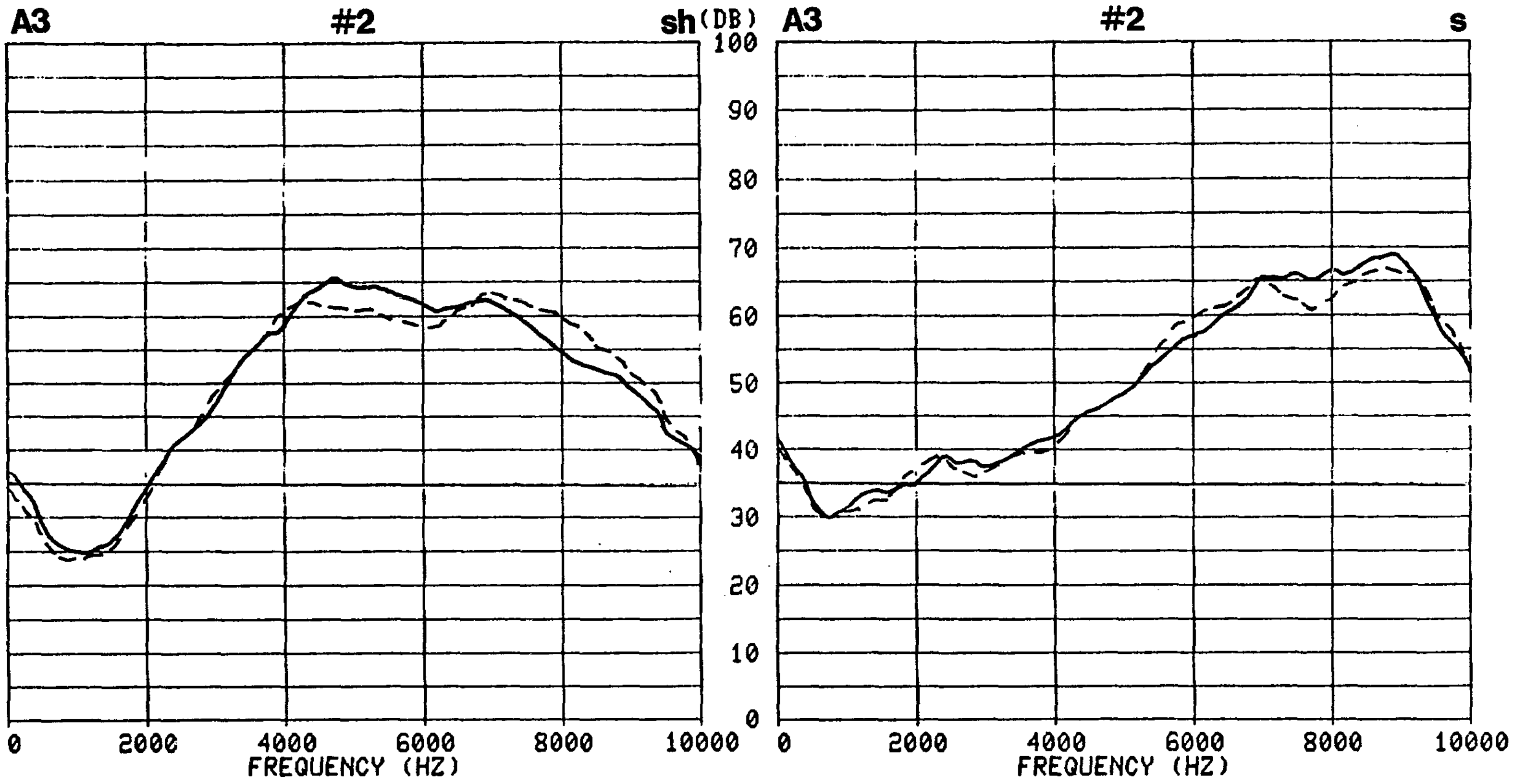
A1
#1
sh
(DB)
A1
#1
s
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

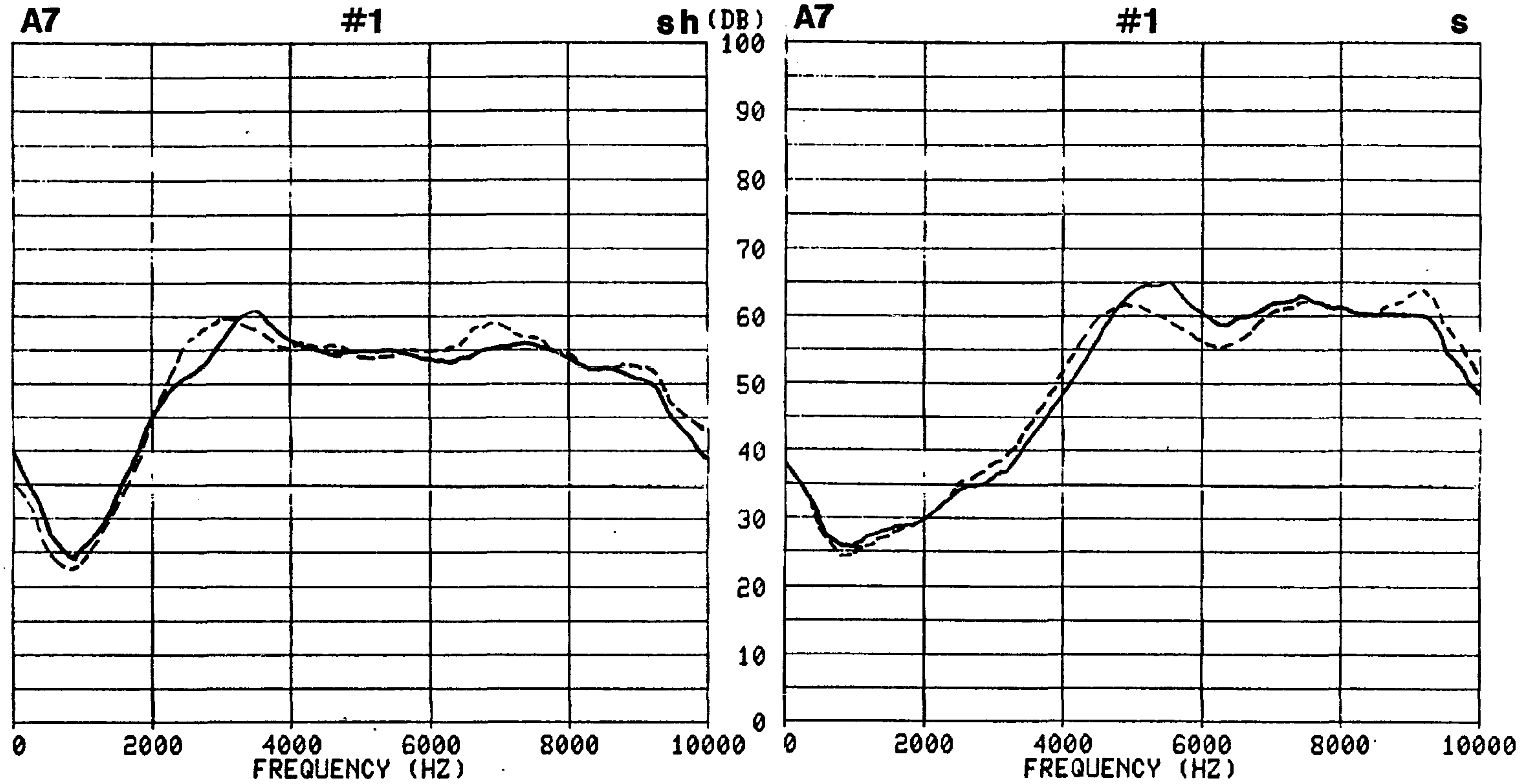

A1
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
A1
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

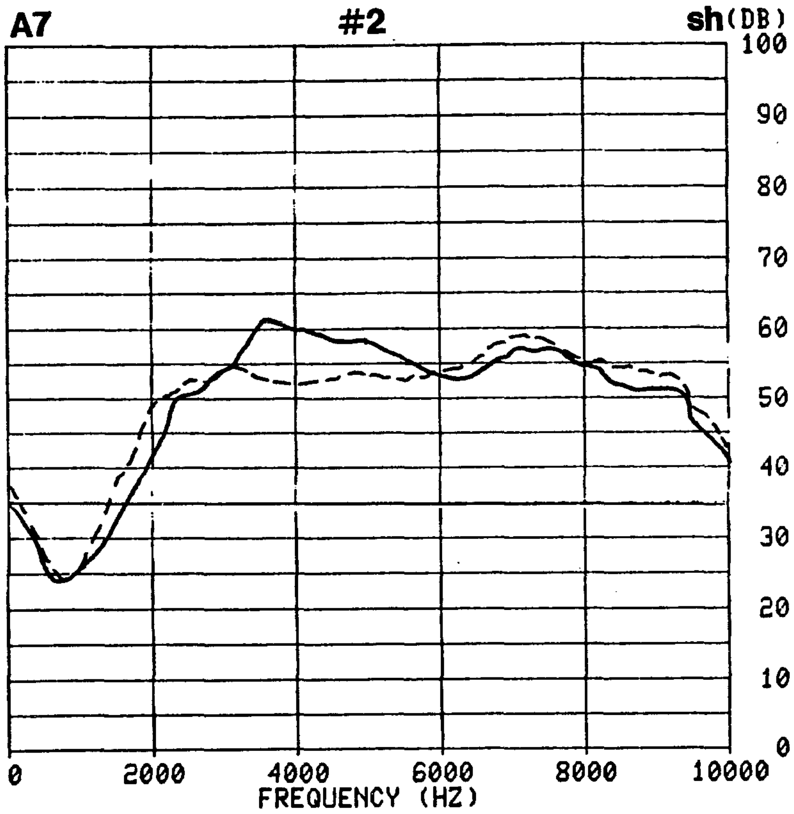

A6
#1
sh
A6
#1
s
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

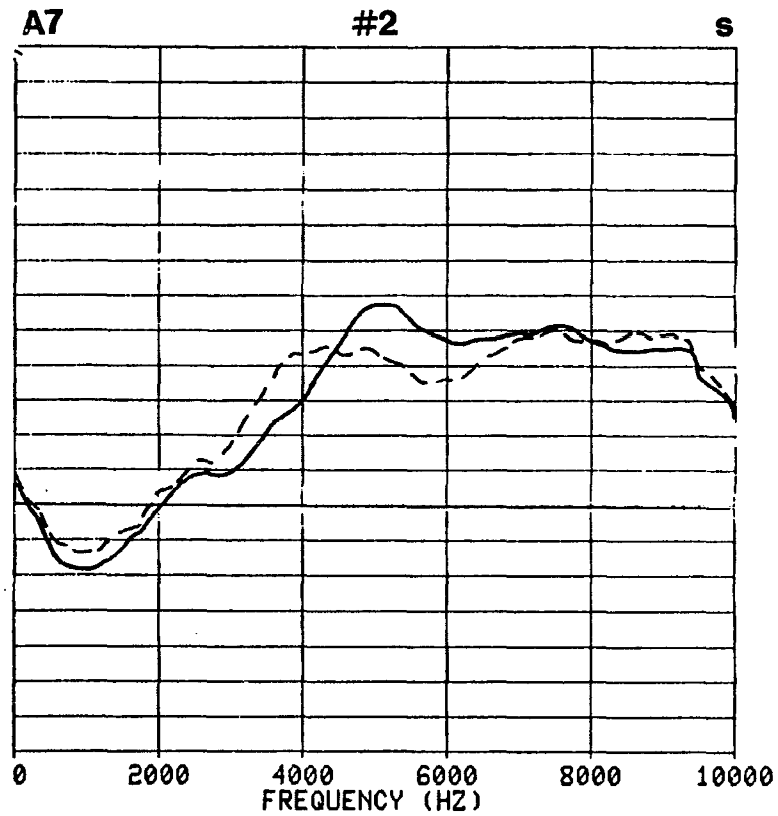

A6
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
A6
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

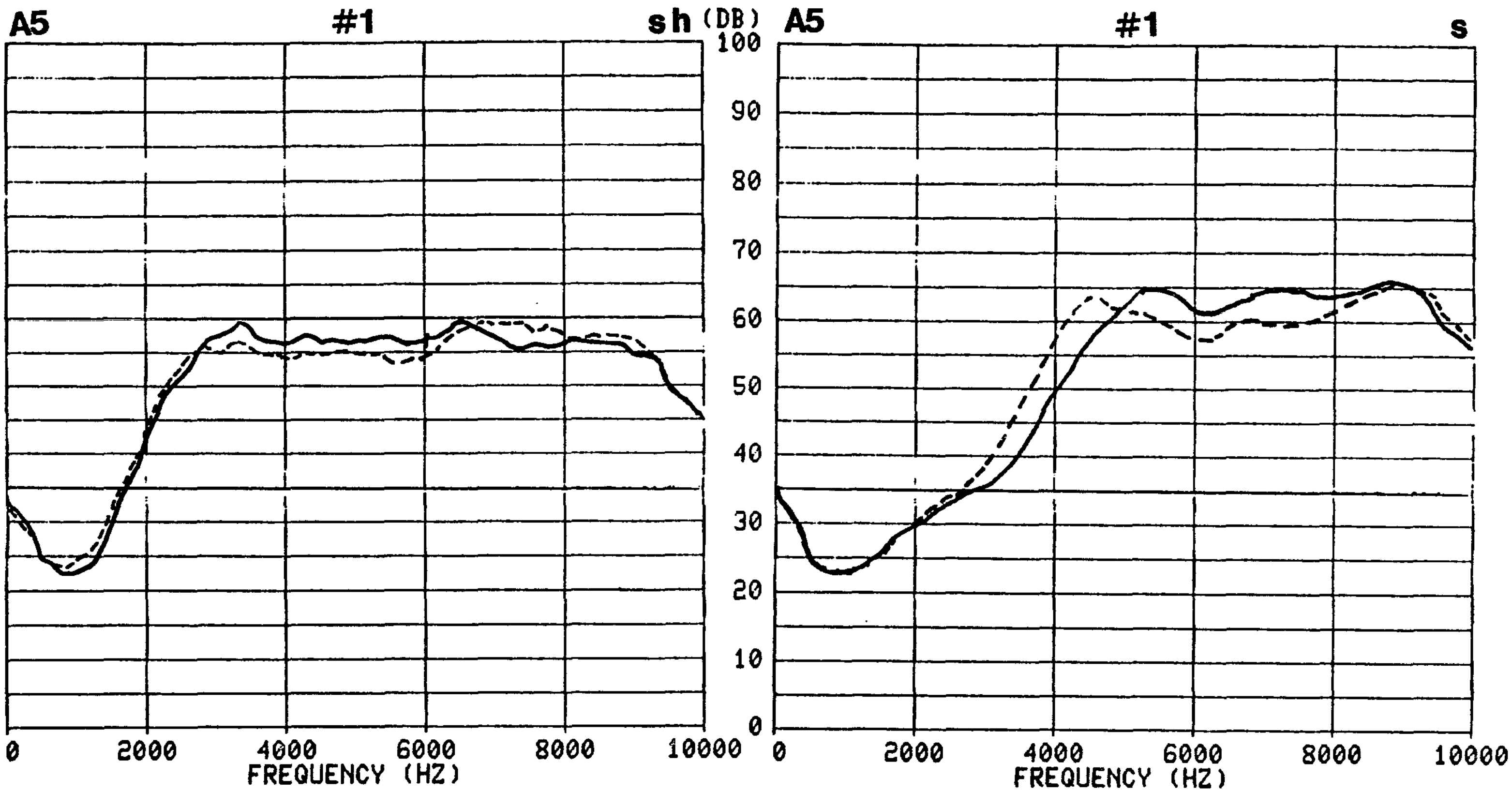

A3
#1
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
A3
#1
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

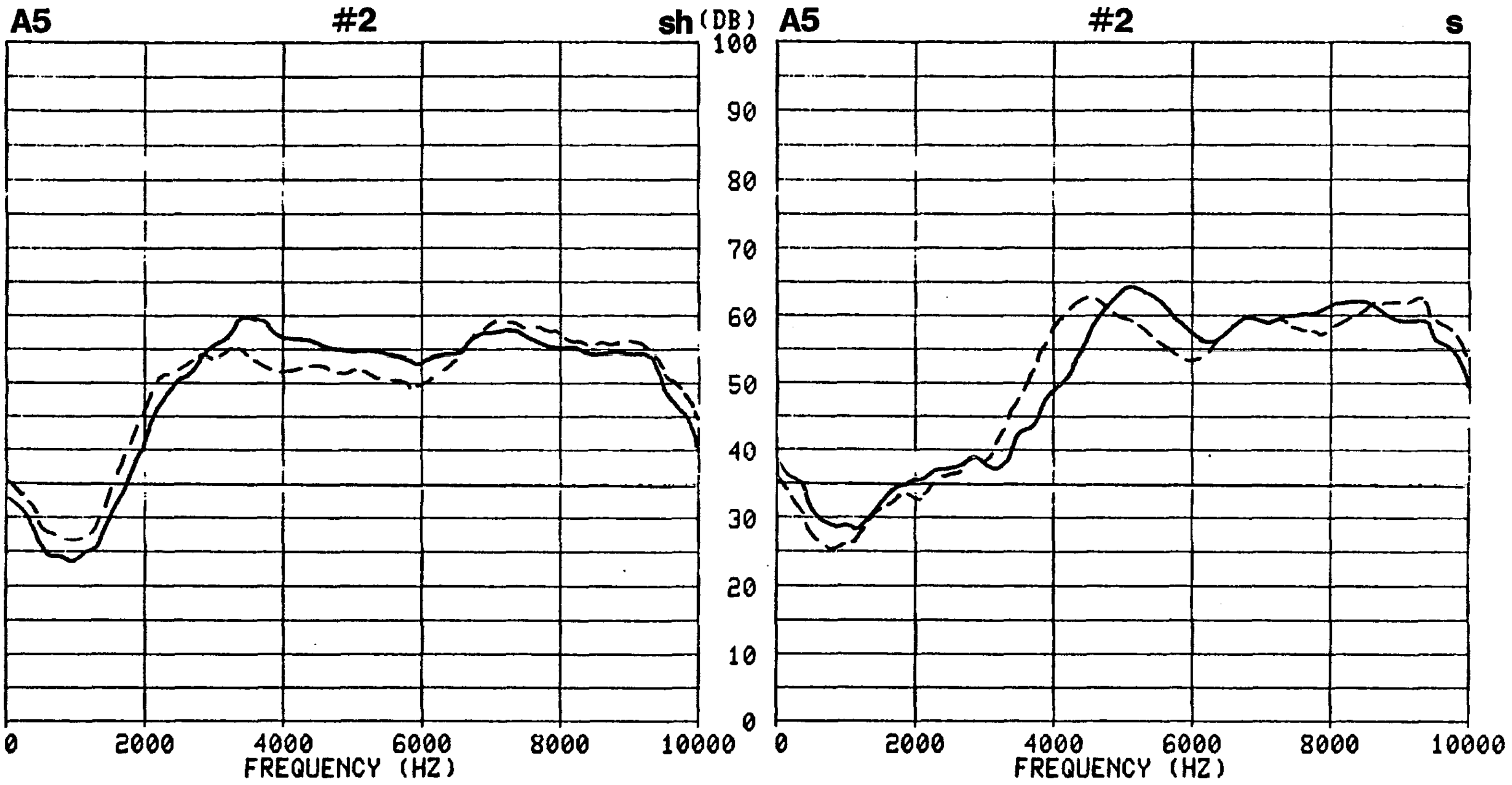

A3
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
A3
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

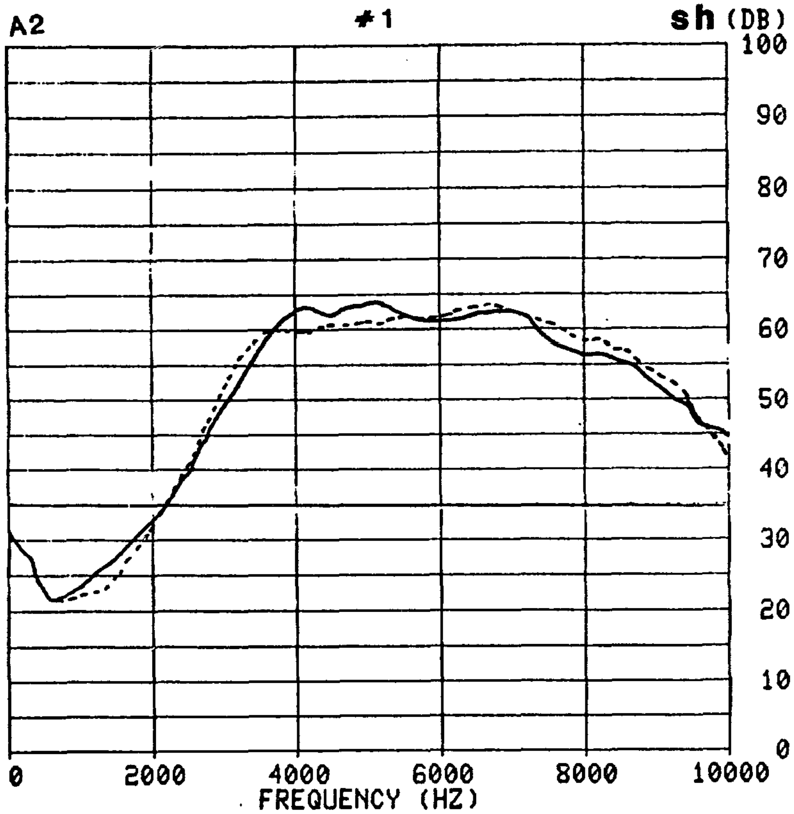

A7
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
A7
#1
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

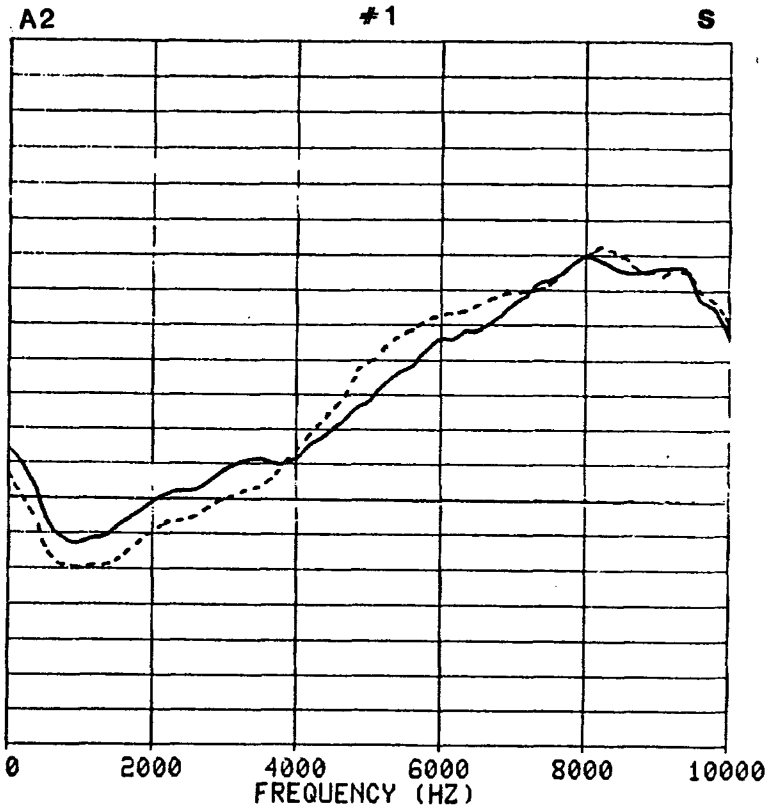

A7
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

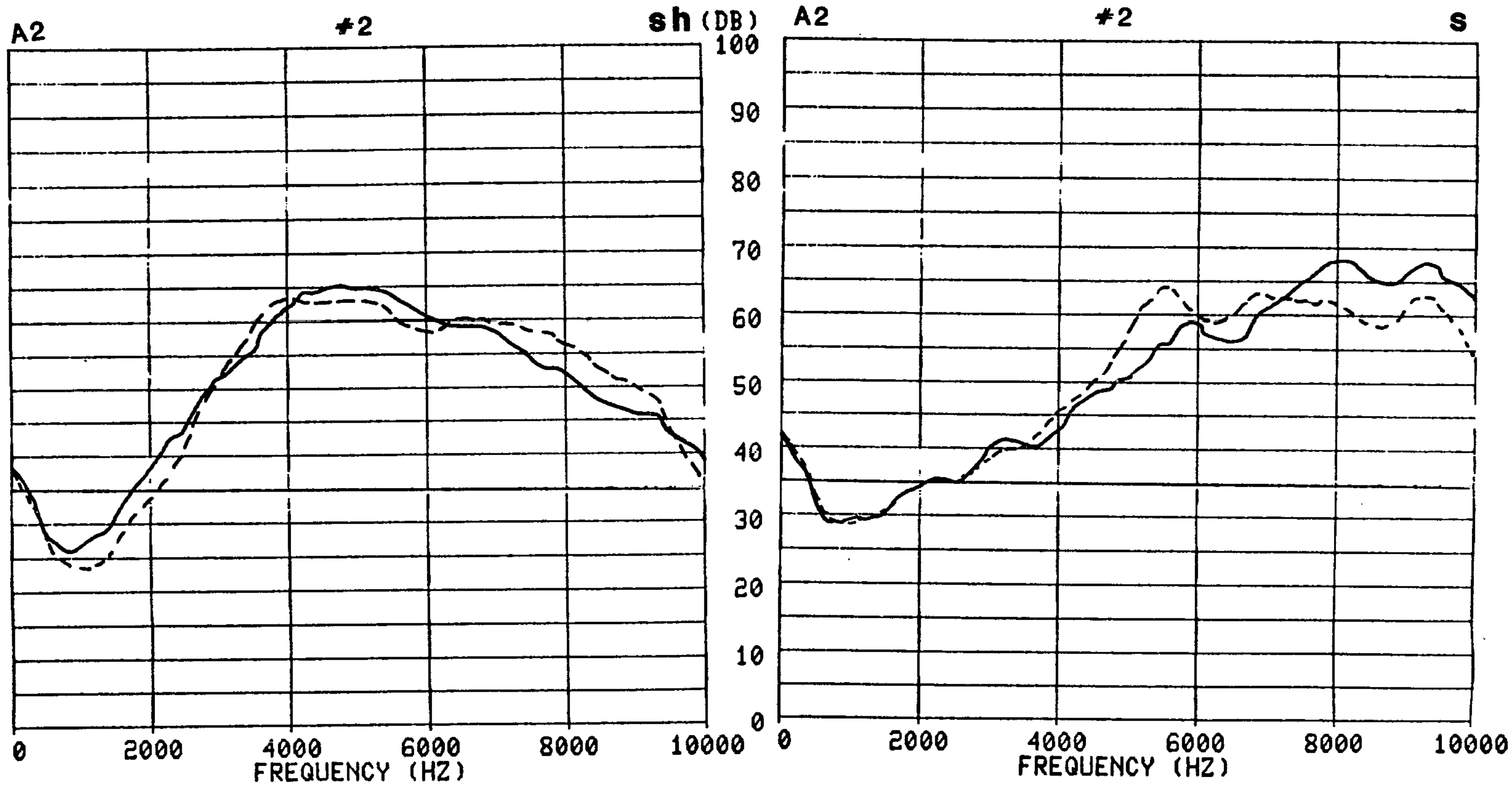

A7
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

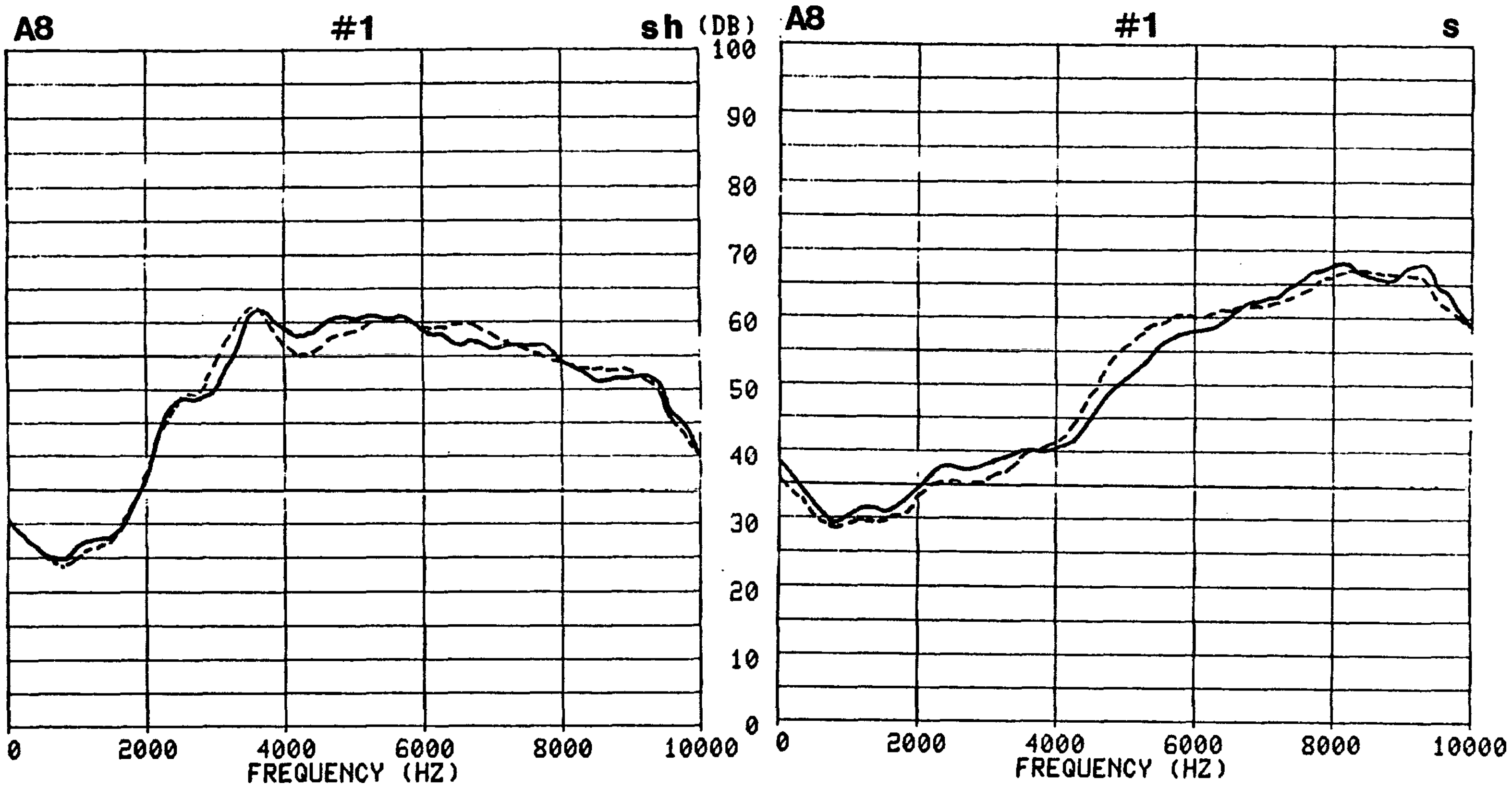

A5
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
A5
#1
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

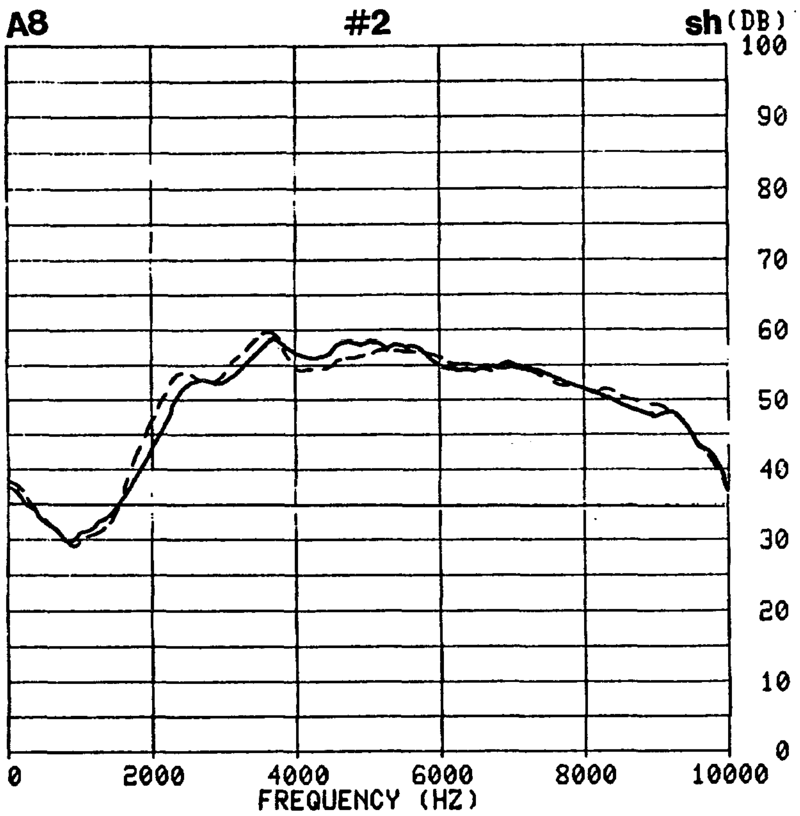

A5
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
A5
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

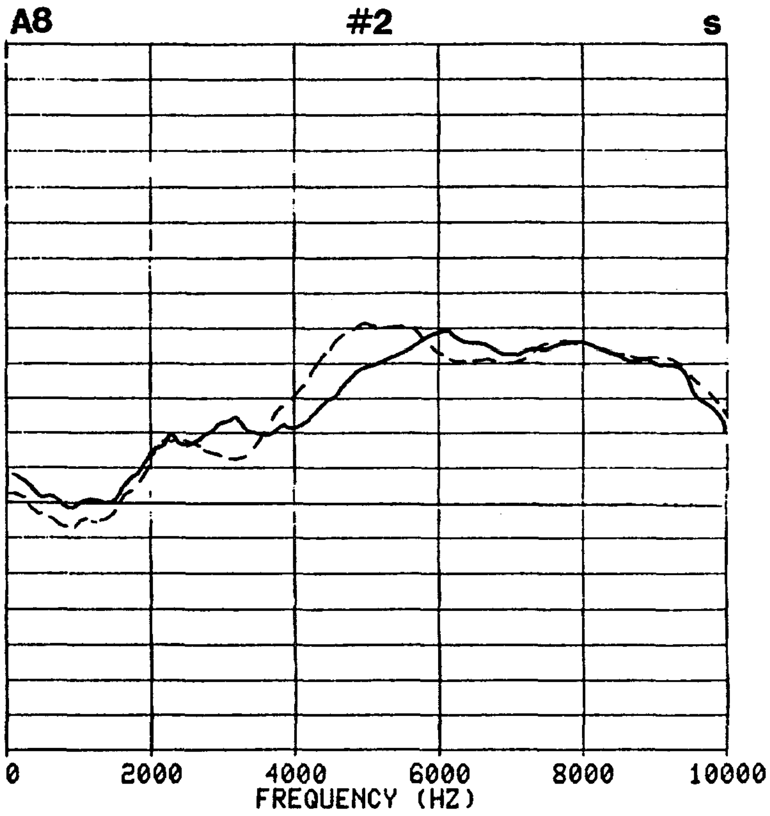

A2
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

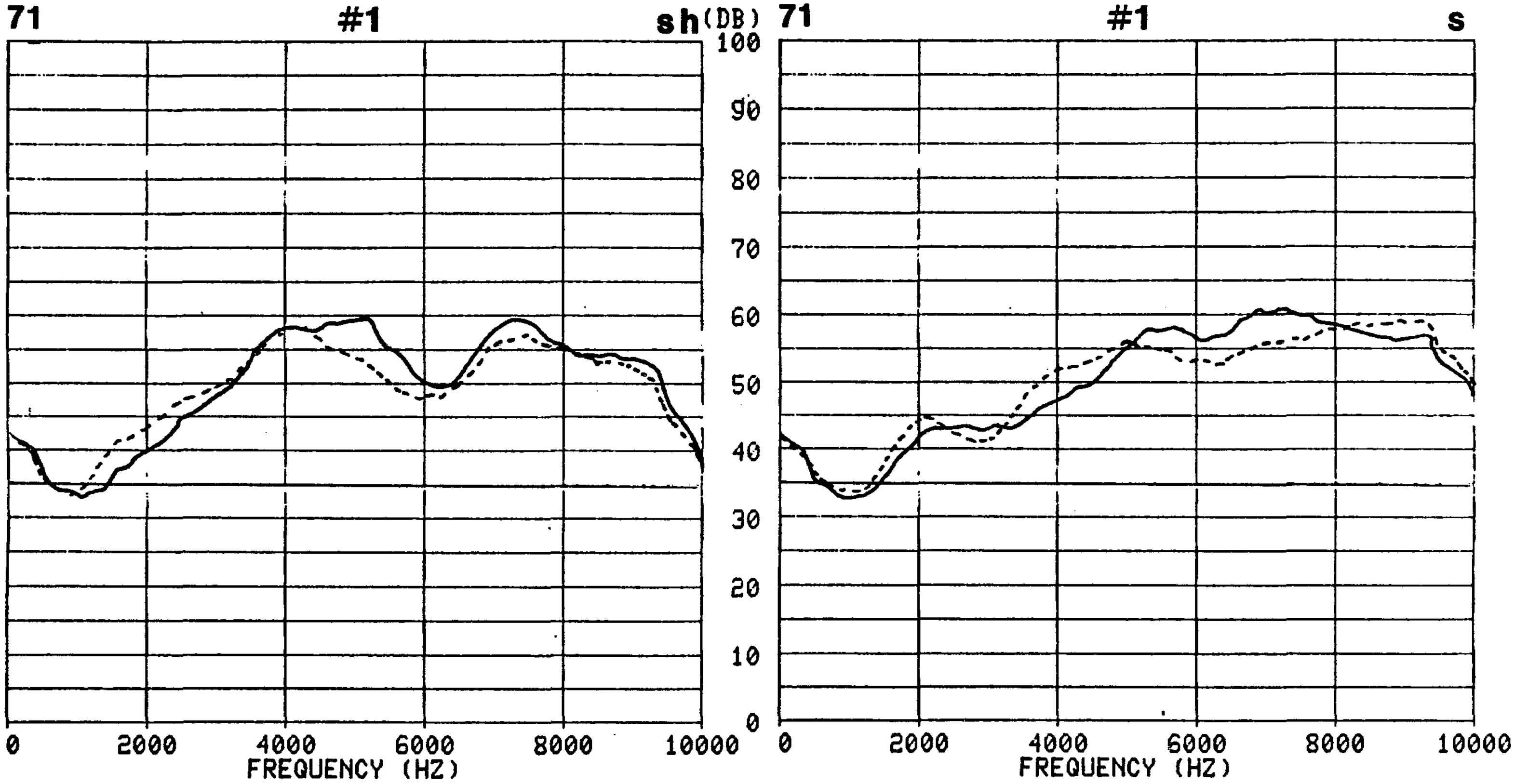

A2
#1
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

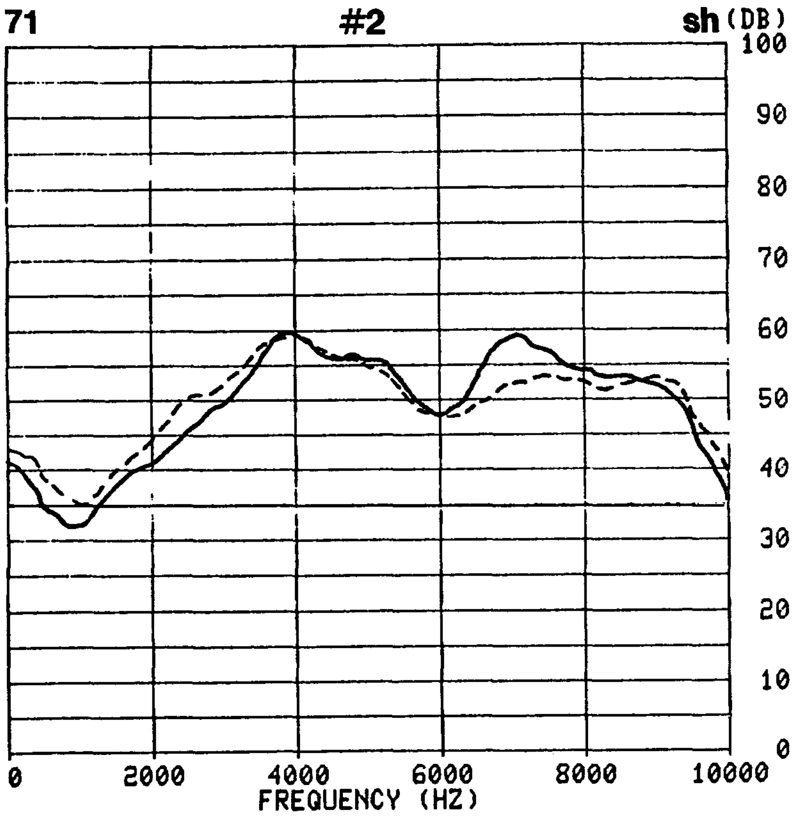

A2
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
A2
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

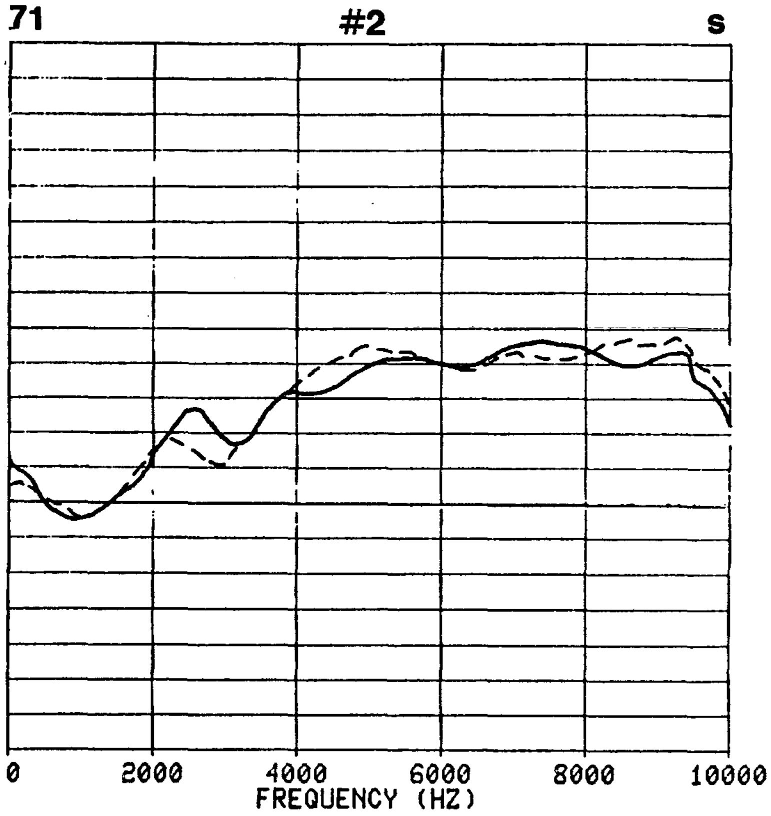

A8
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
A8
#1
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

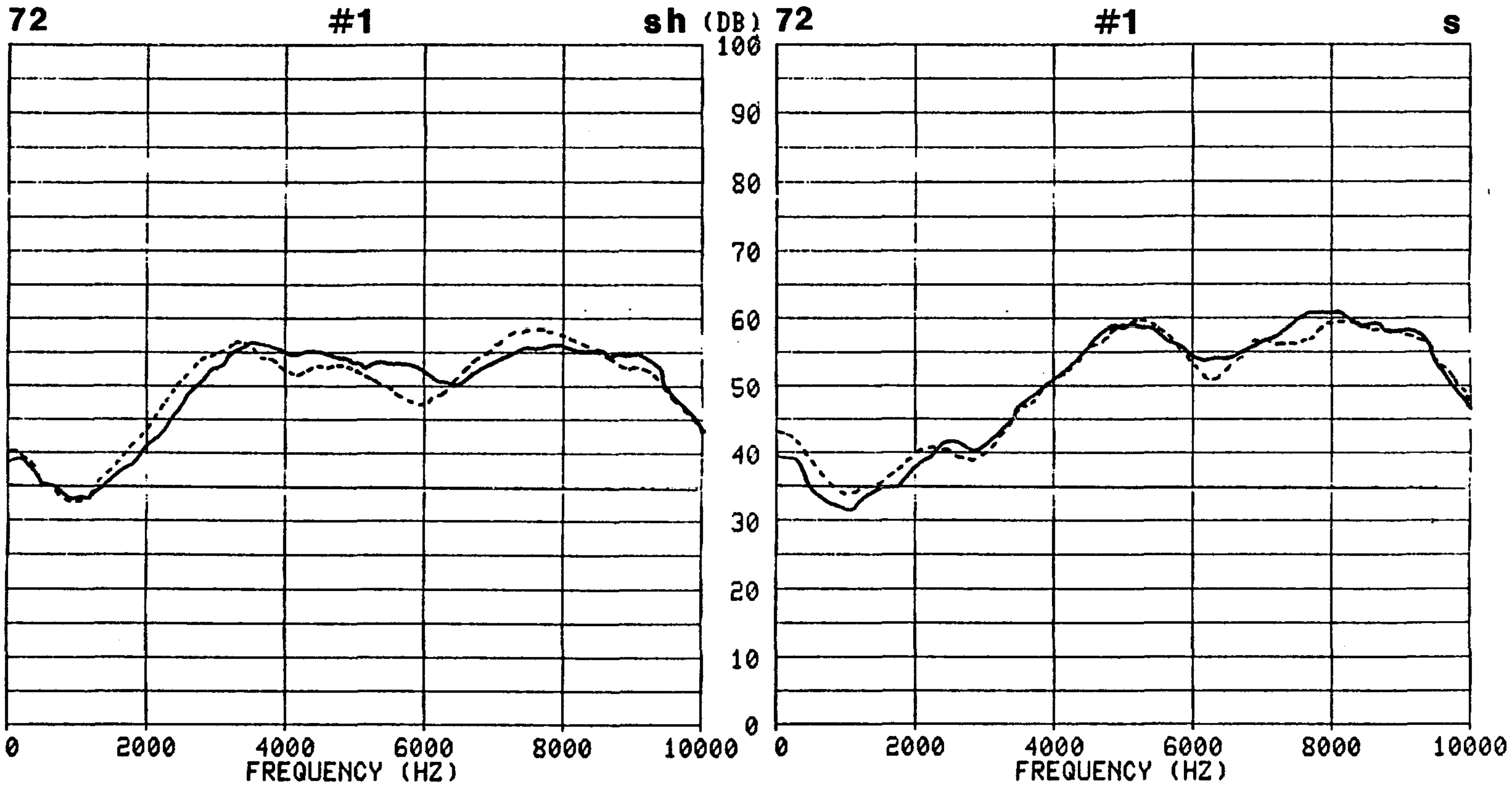

A8
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

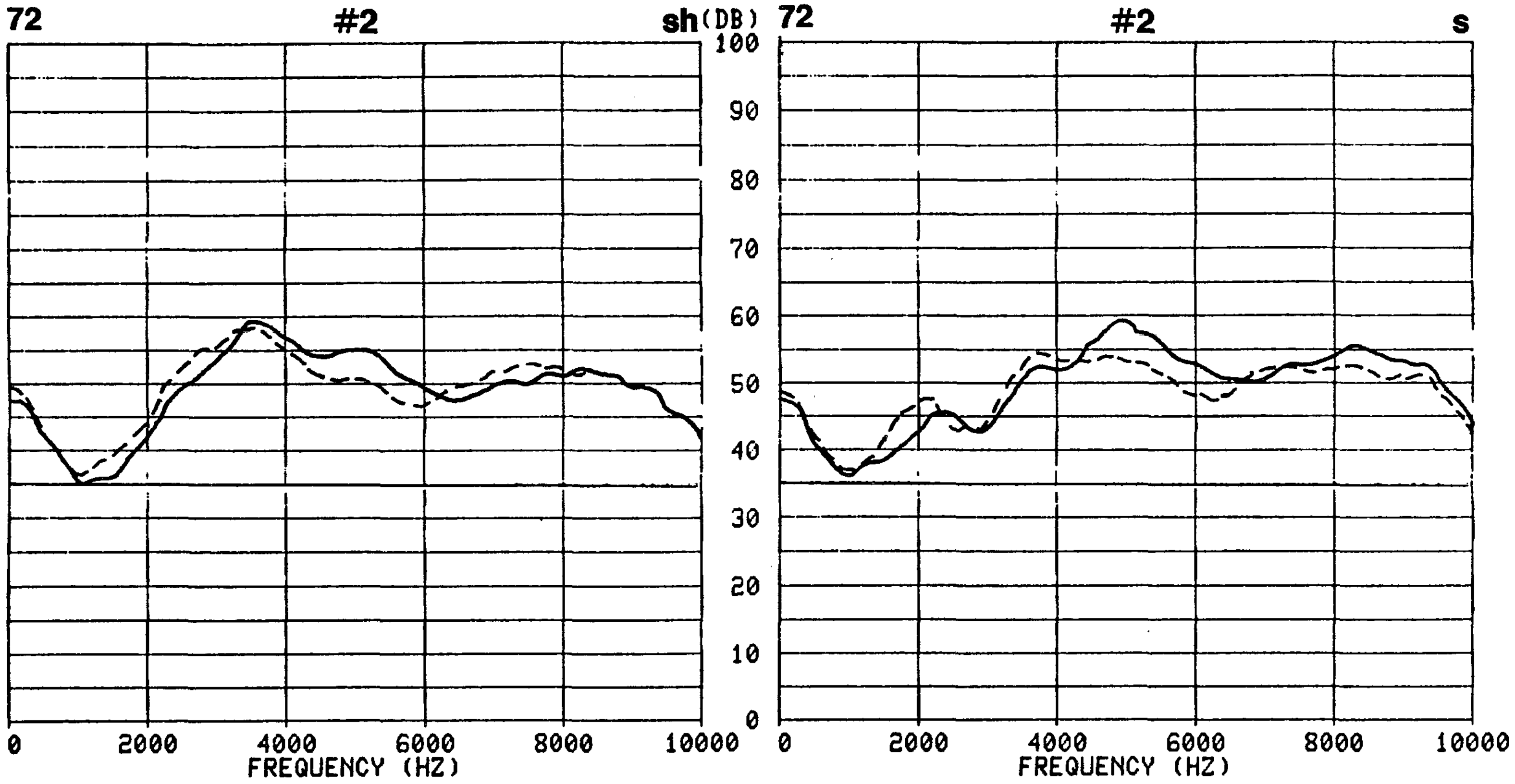

A8
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

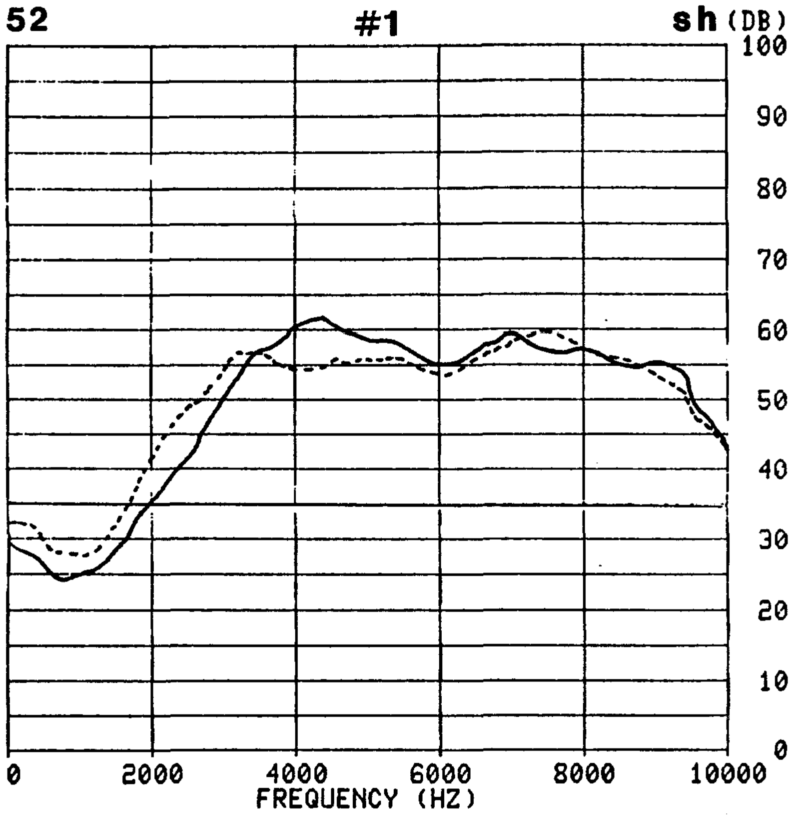

71
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
71
#1
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

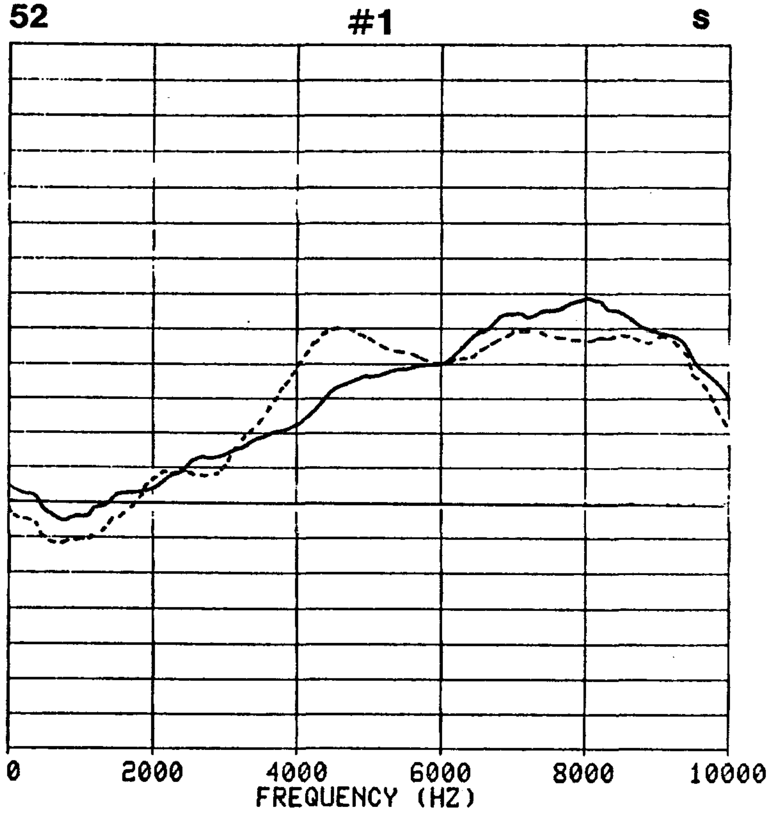

71
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

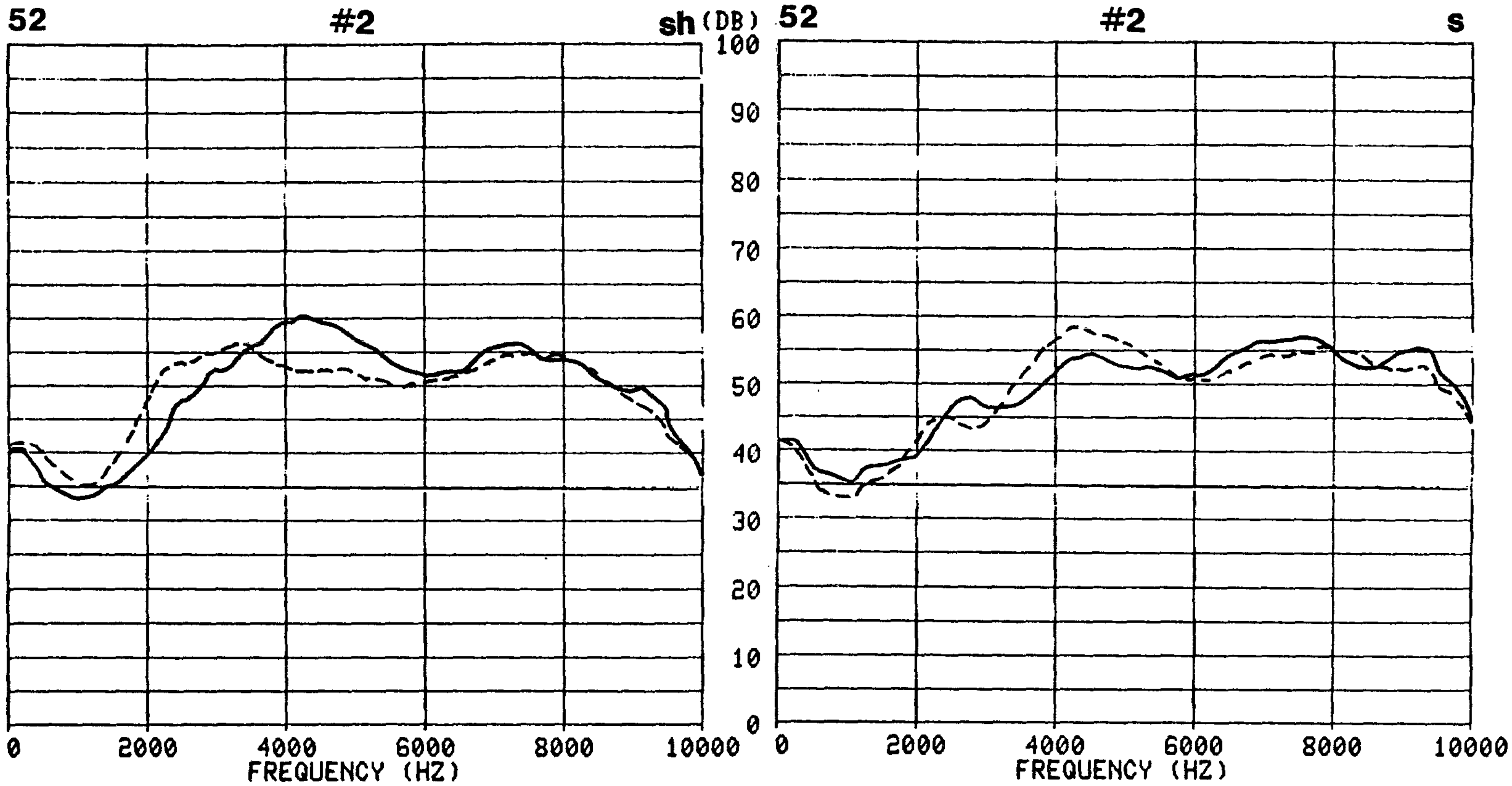

71
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

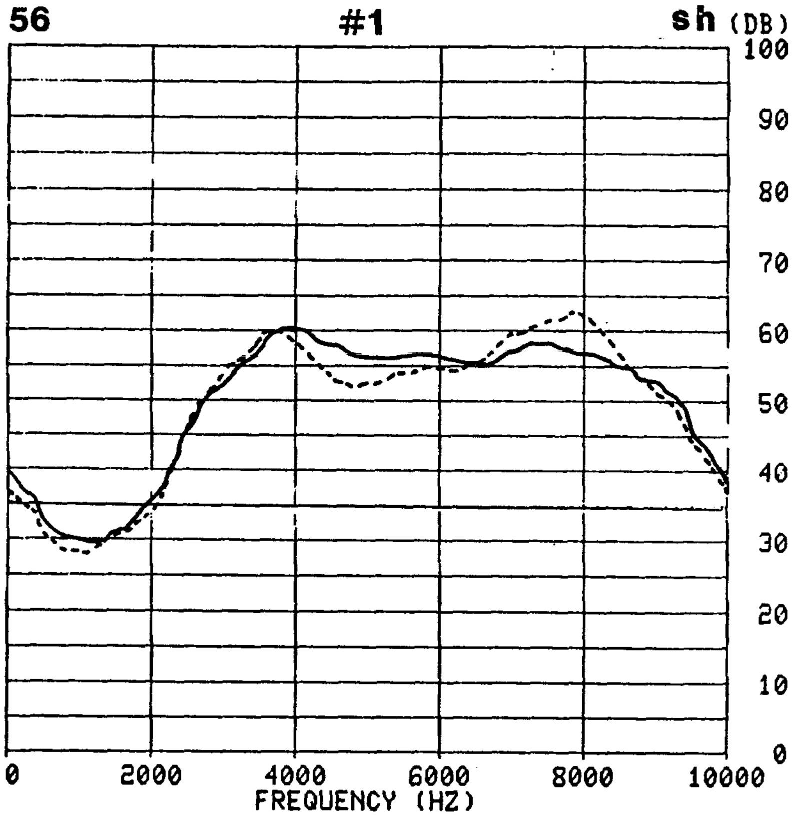

72
#1
sh
(DB)
72
#1
s
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

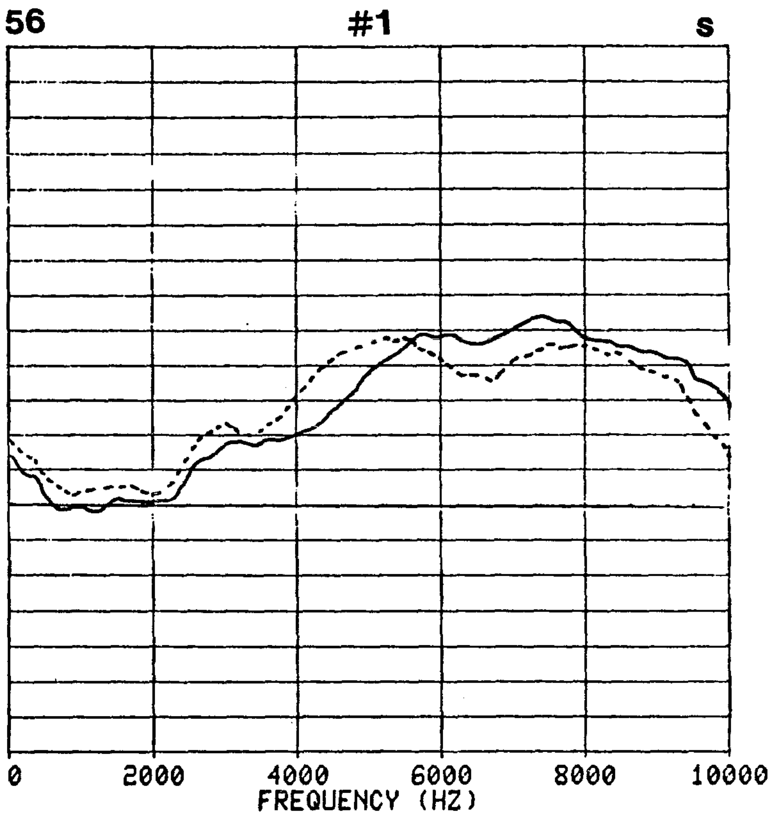

72
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
72
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

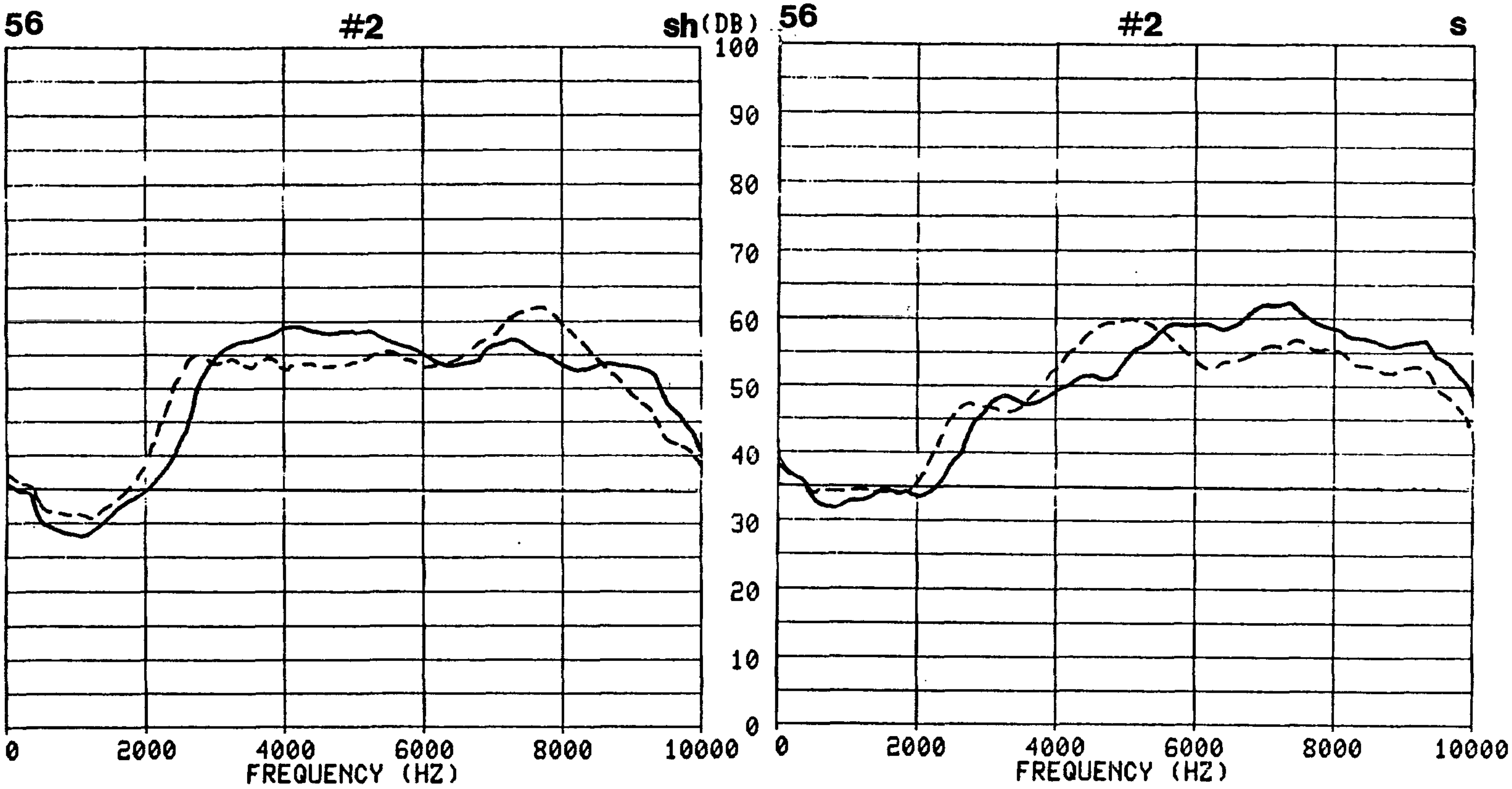

52
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

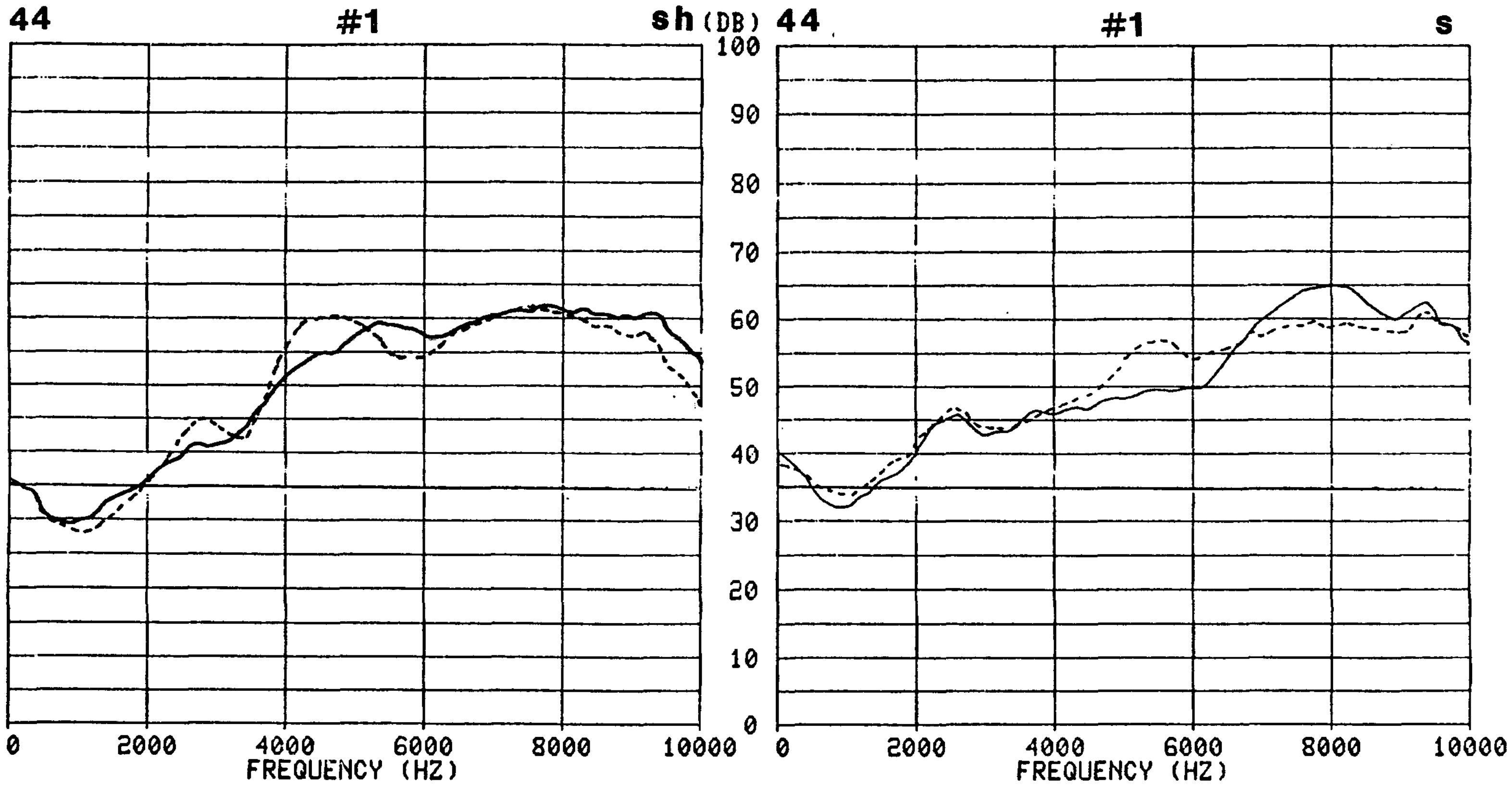

52
#1
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

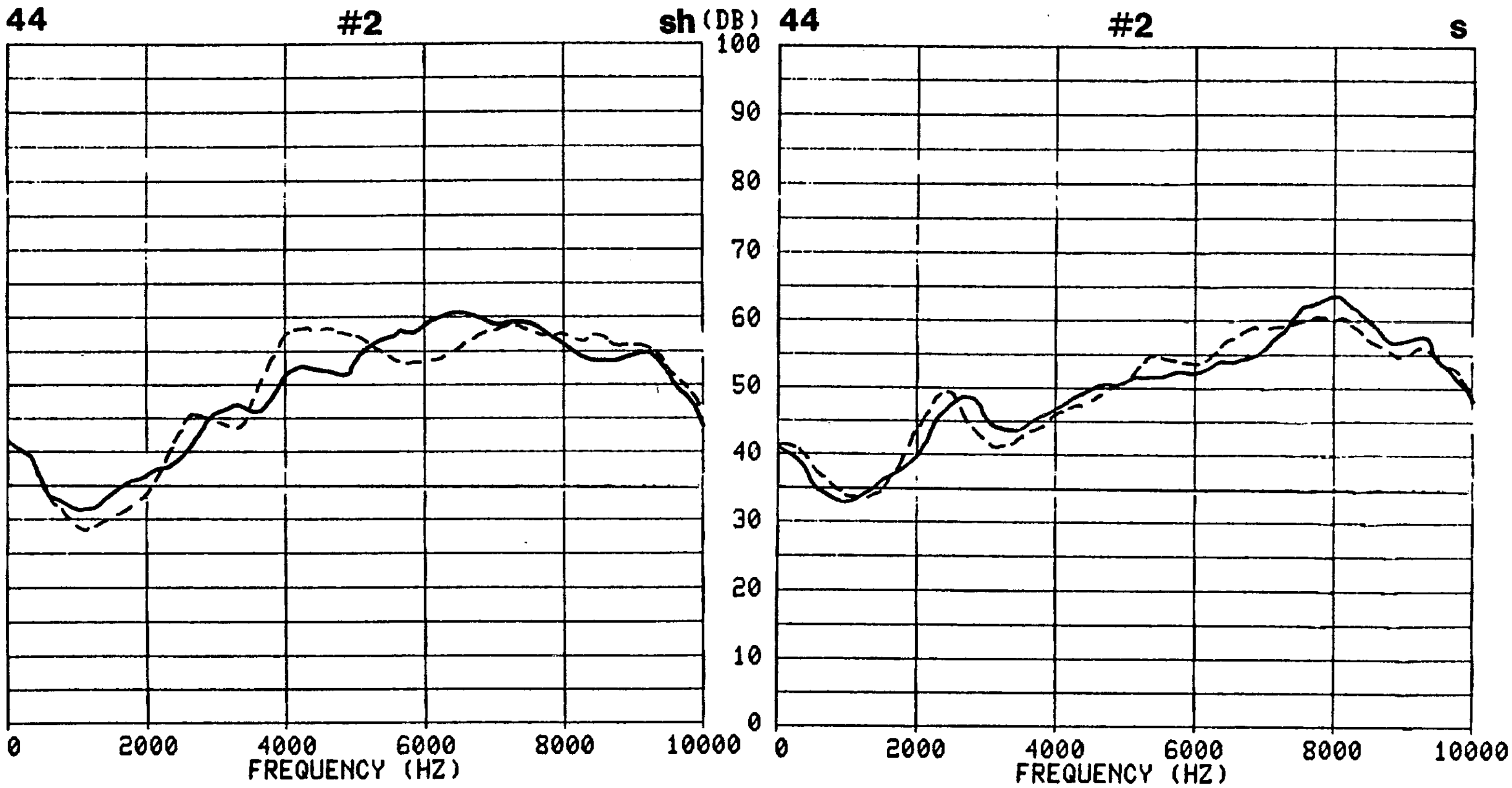

52
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
52
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

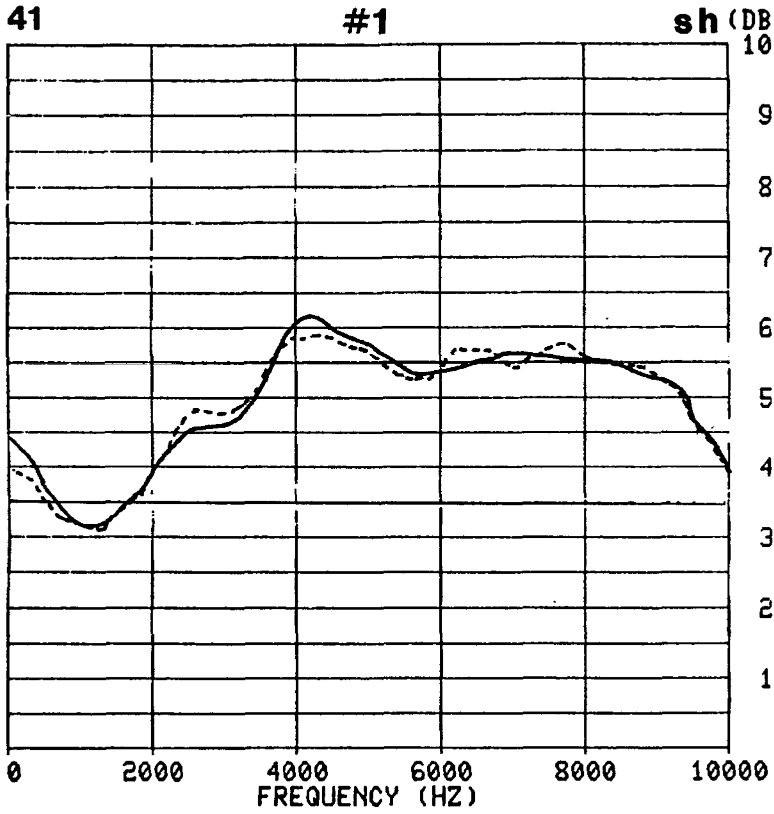

56
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

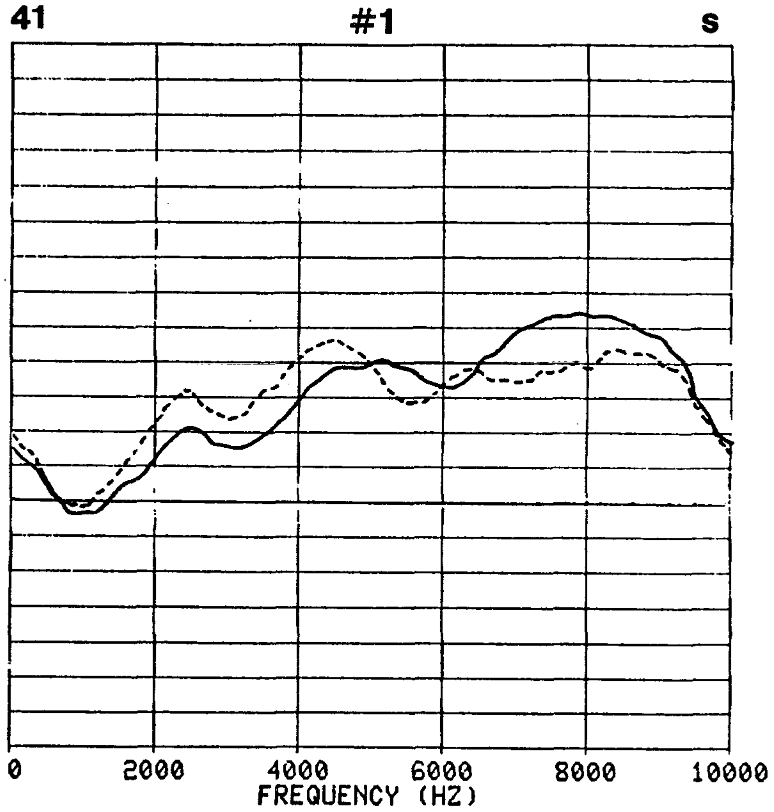

56
#1
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

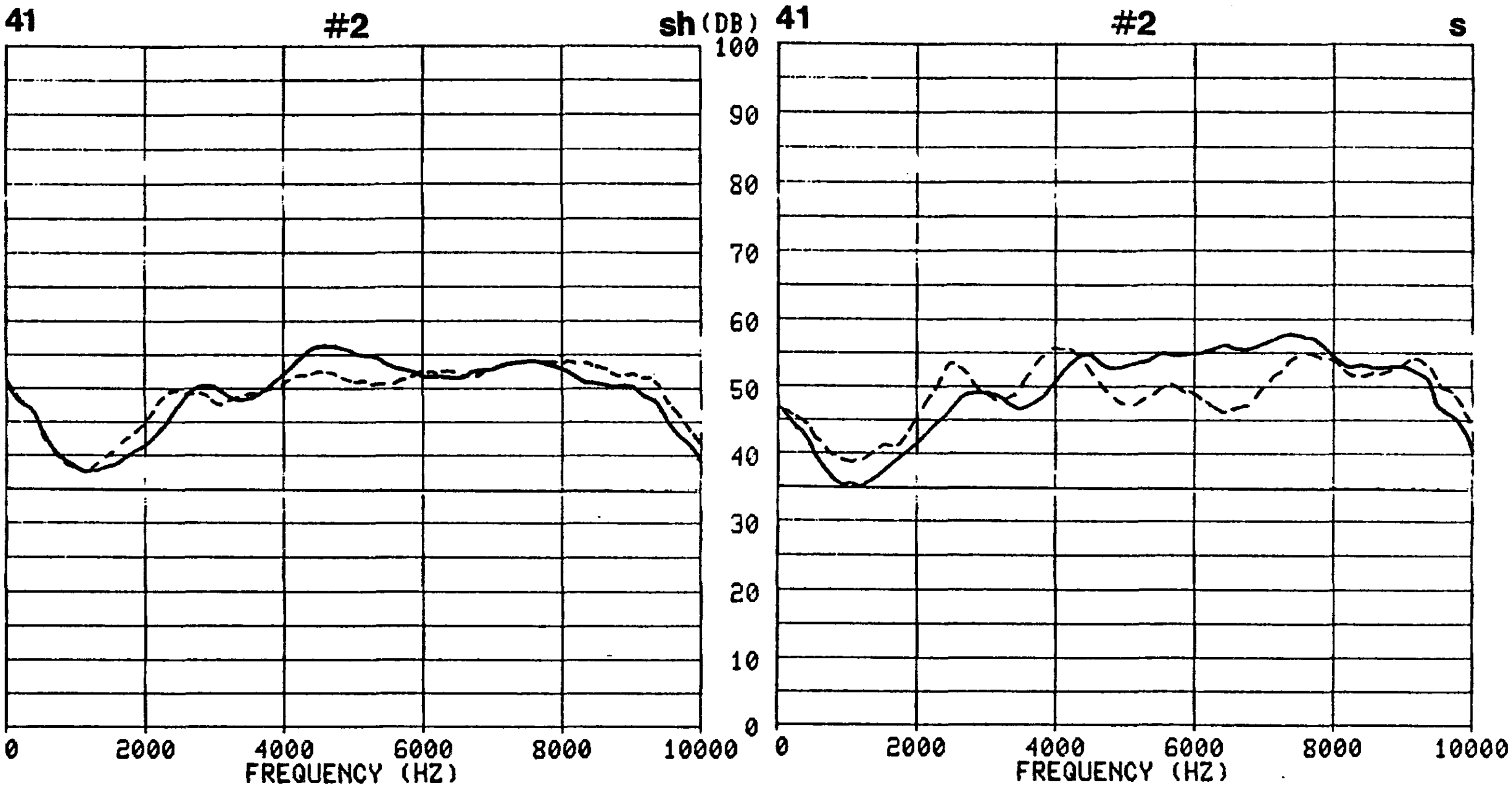

56
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
56
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)


44
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
44
#1
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

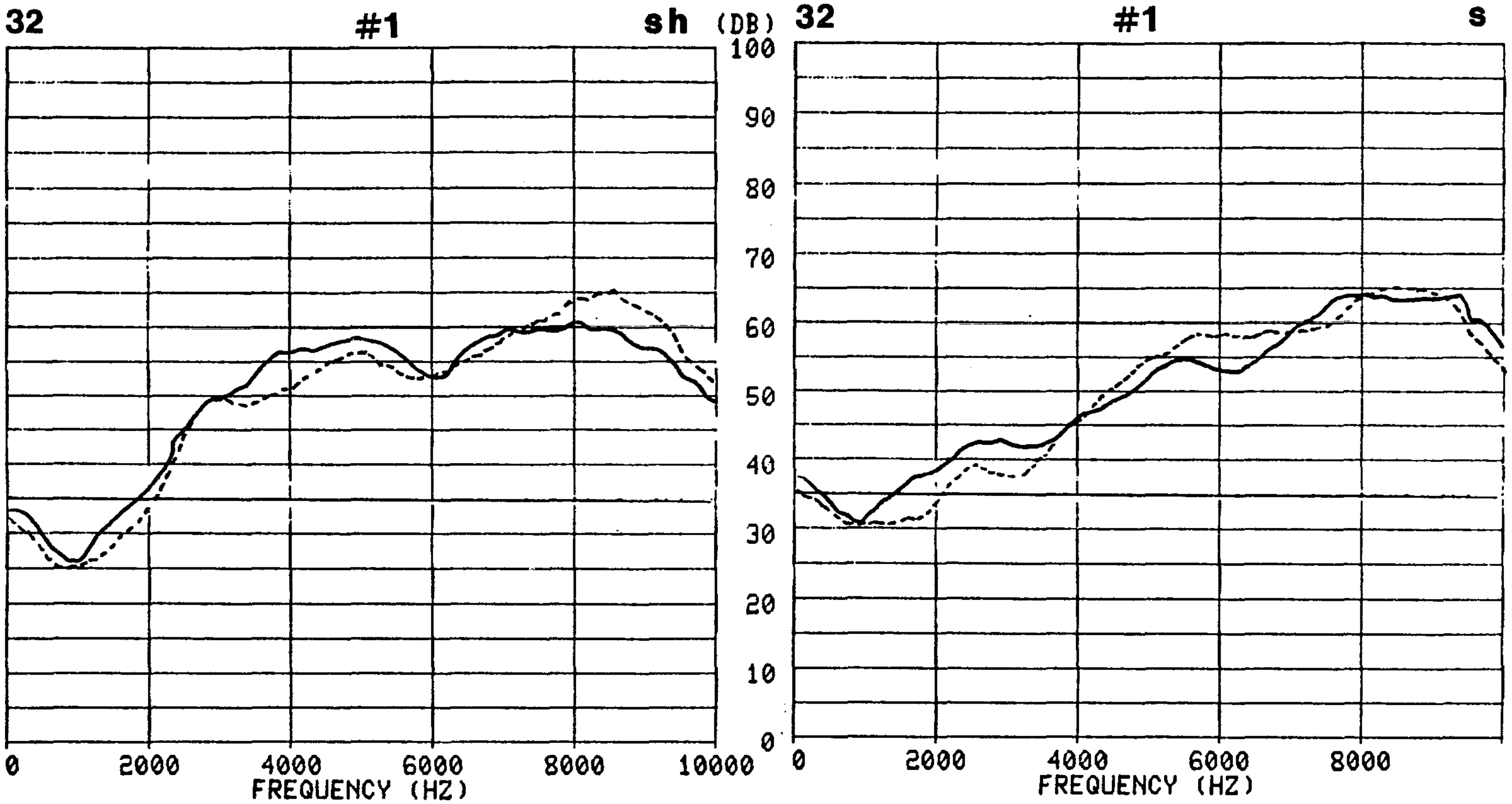

44
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
44
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

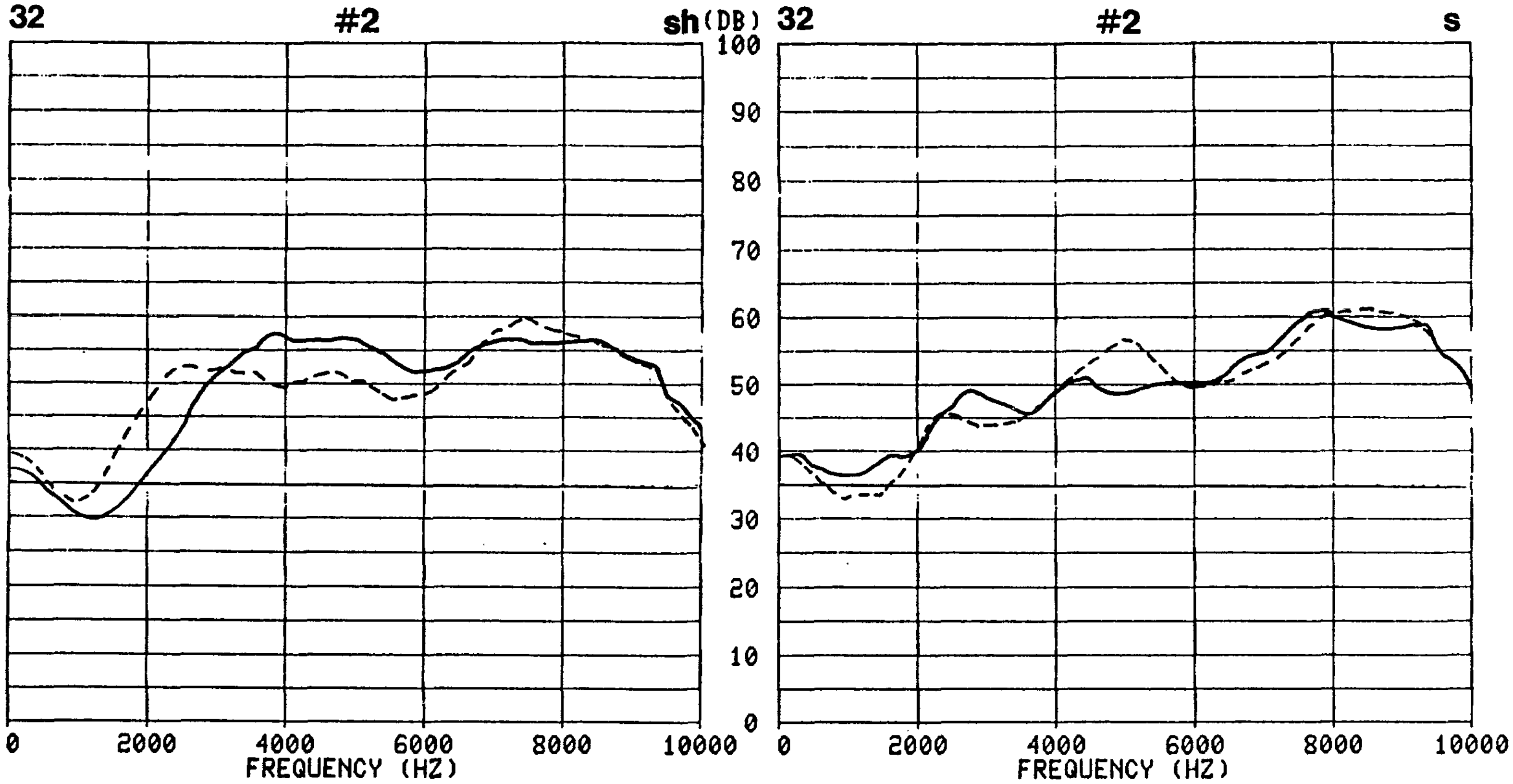

41
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

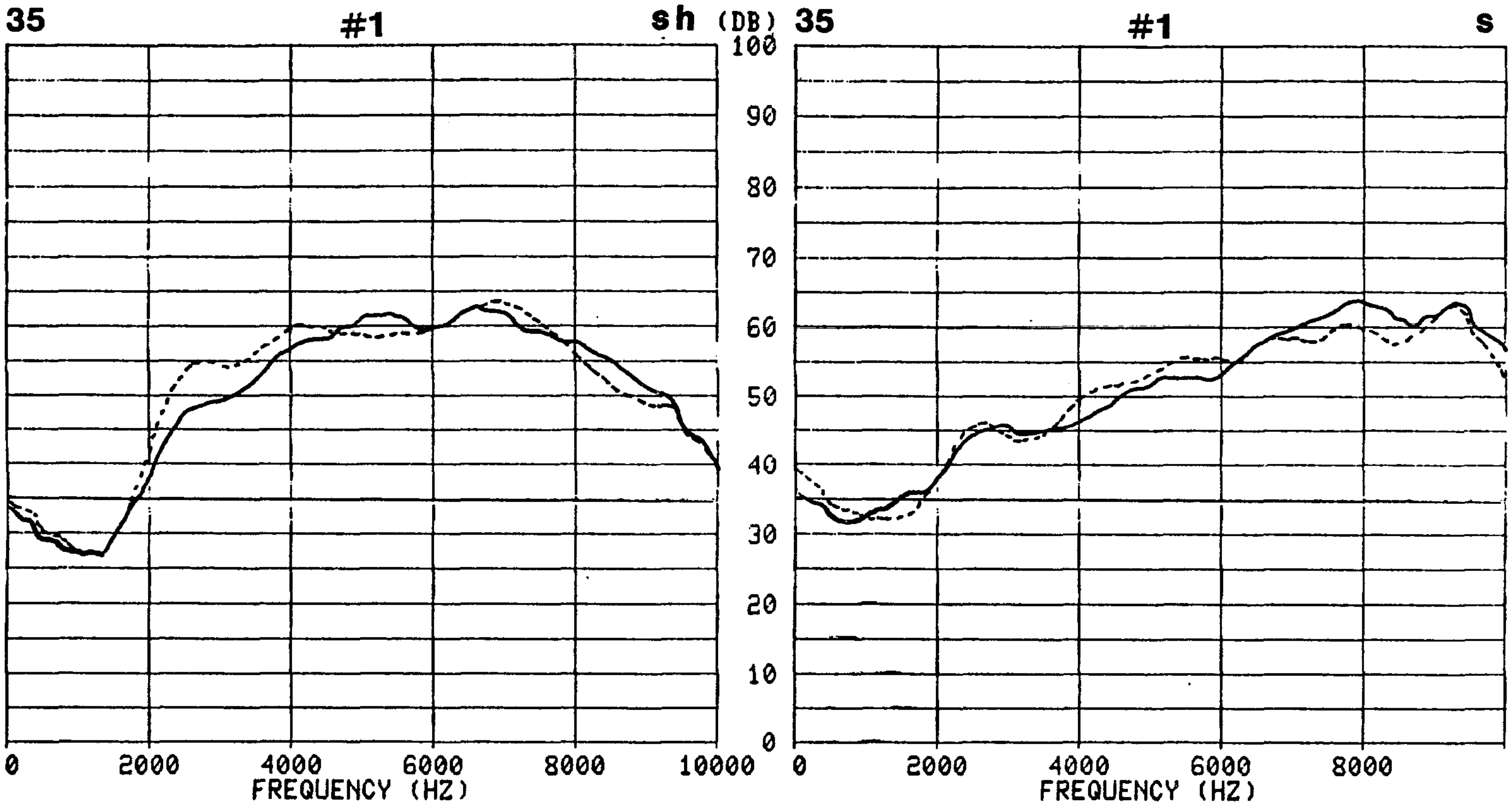

41
#1
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

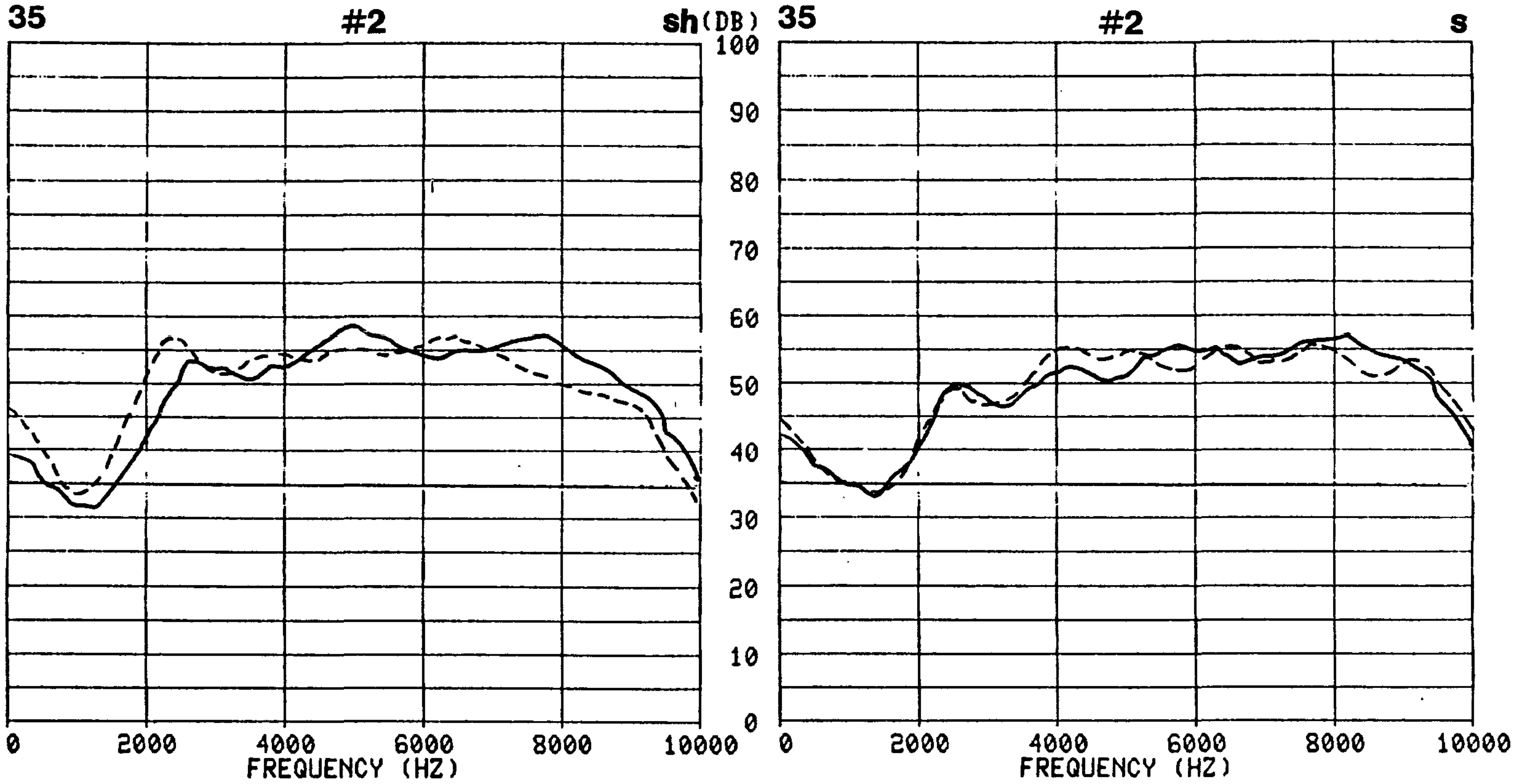


41
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
41
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)


32
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
32
#1
s
0
2000
4000
6000
8000
FREQUENCY (HZ)

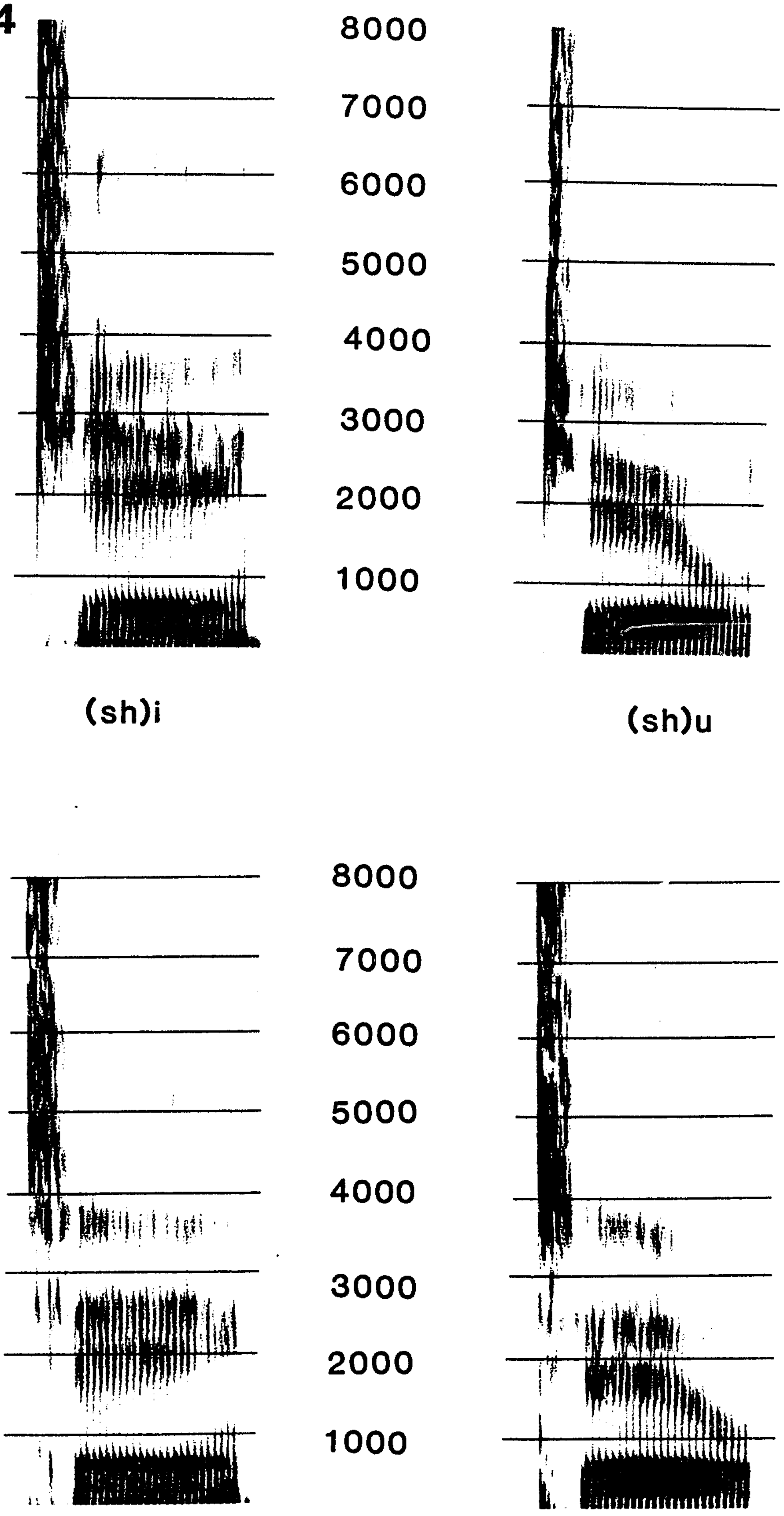

32
#2
sh
(DB)
32
#2
s
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)


35
#1
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
35
#1
s
0
2000
4000
6000
8000
FREQUENCY (HZ)


35
#2
sh
(DB)
100
90
80
70
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)
35
#2
s
0
2000
4000
6000
8000
10000
FREQUENCY (HZ)

Appendix F: Sample spectrograms

A4



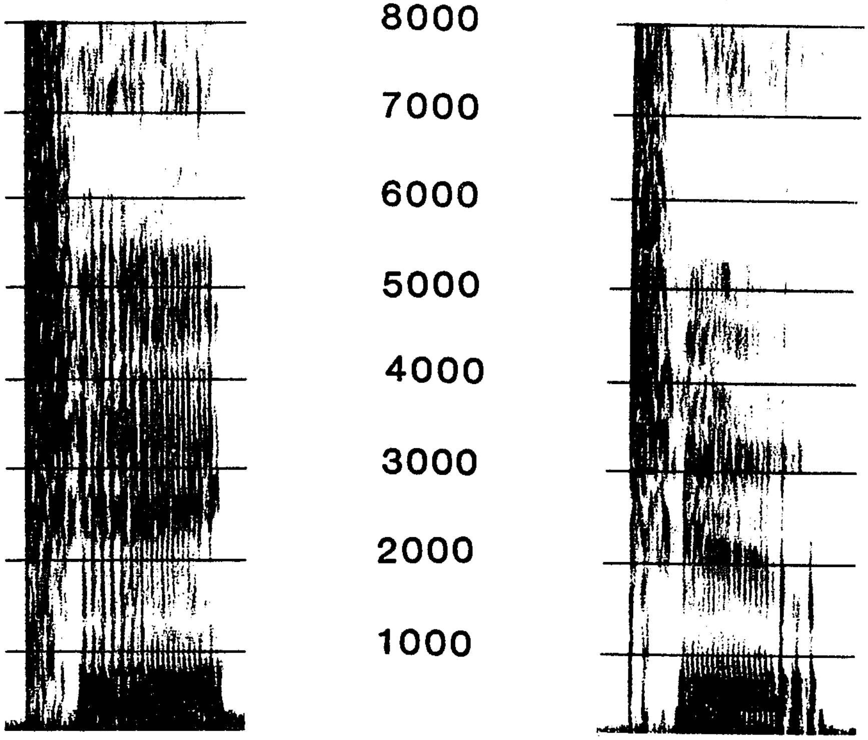
(sh)i

(sh)u

(s)i

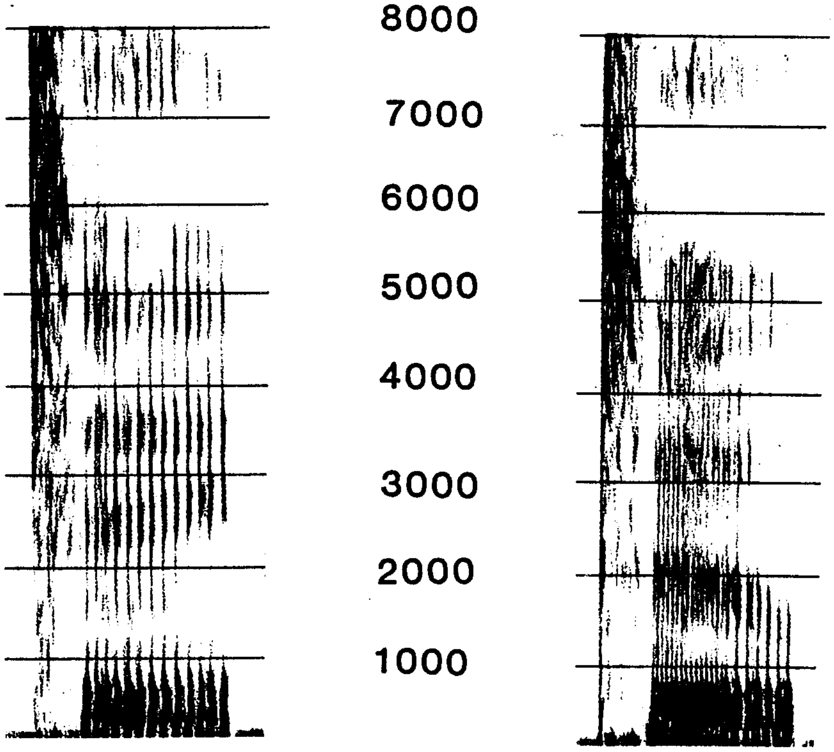
(s)u

A1



(sh)i

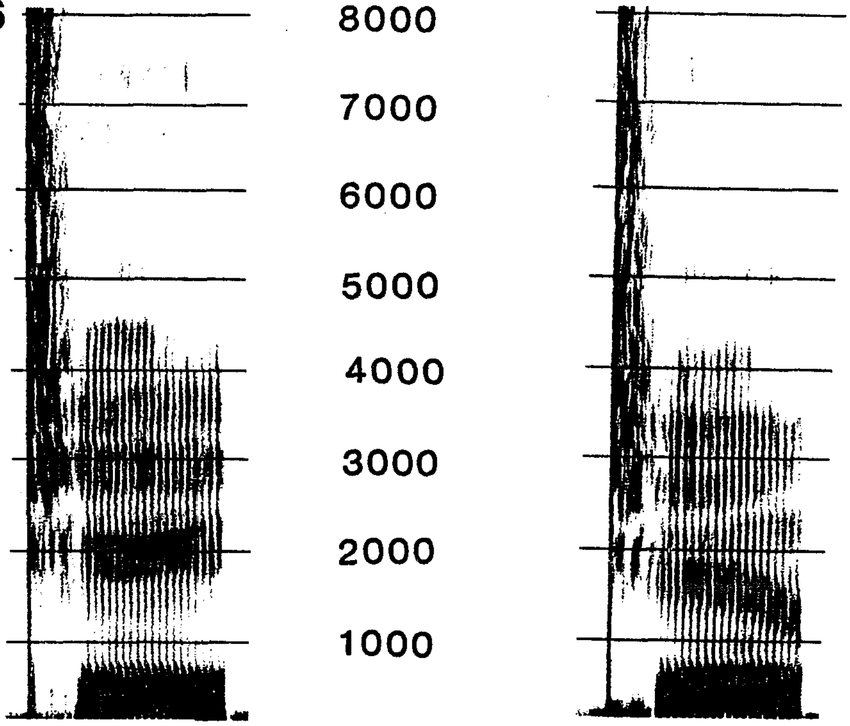
(sh)u



(s)i

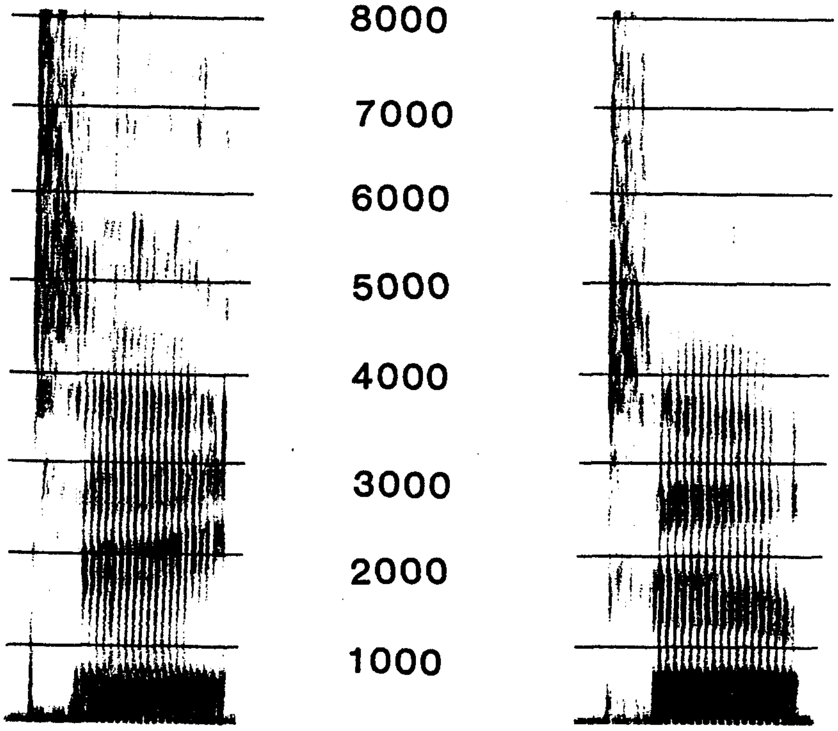
(s)u

A6



(sh)i

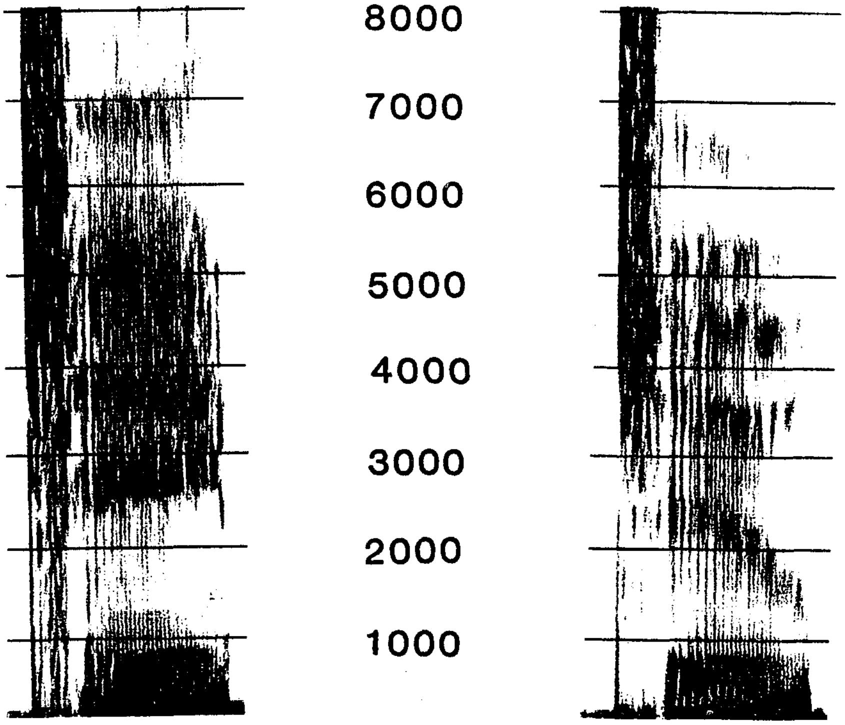
(sh)u



(s)i

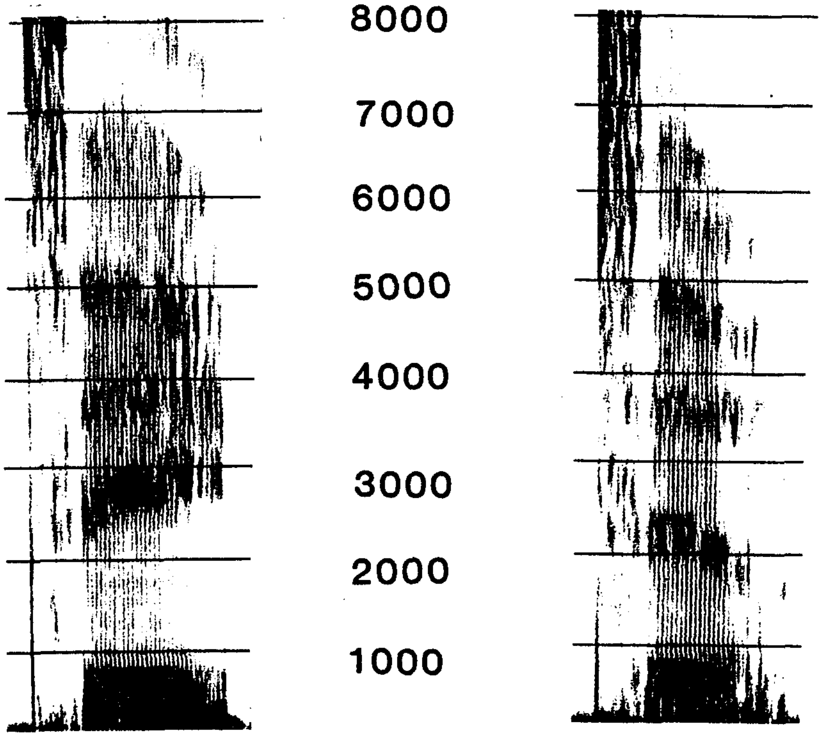
(s)u

A3



(sh)i

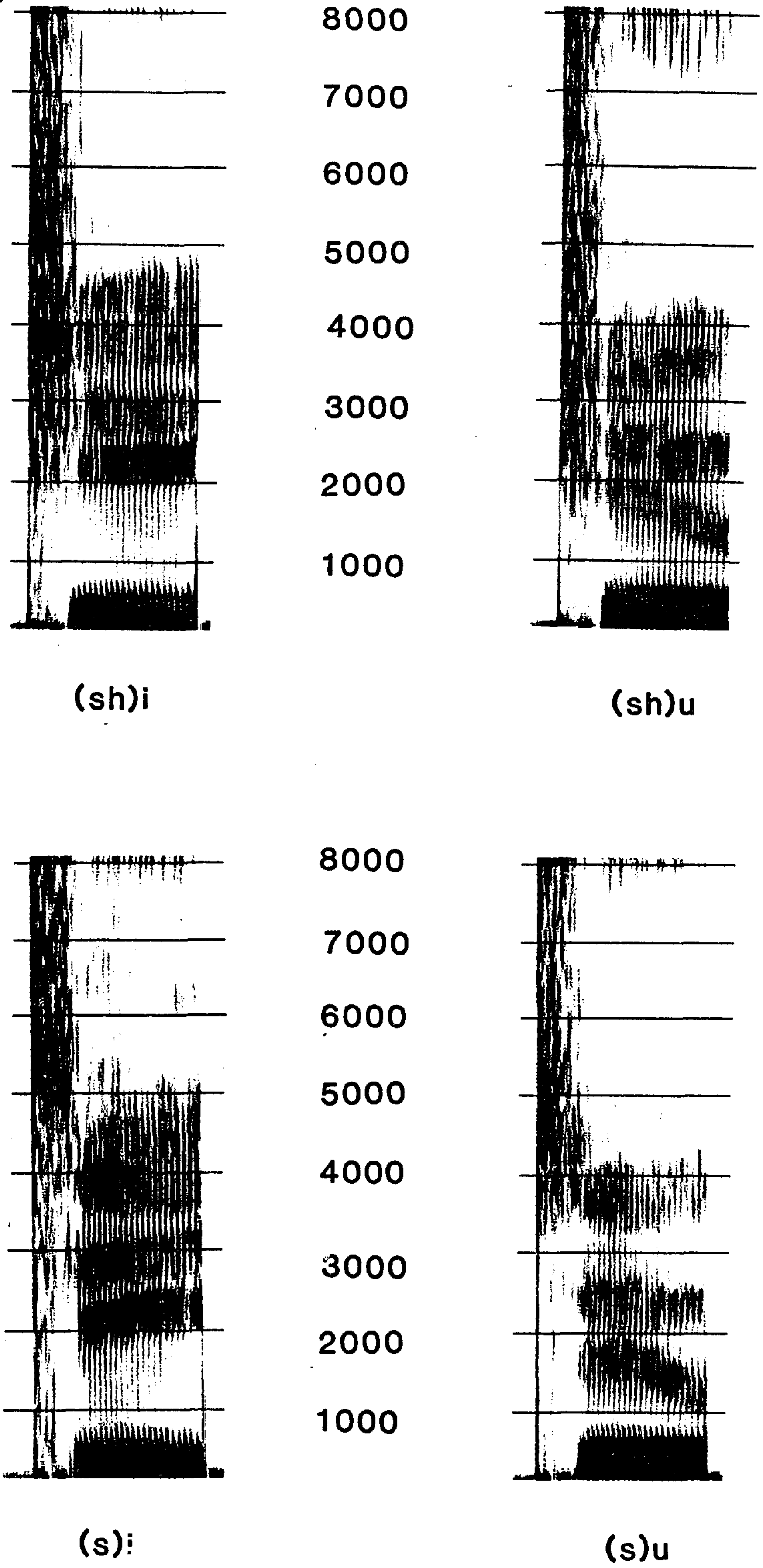
(sh)u



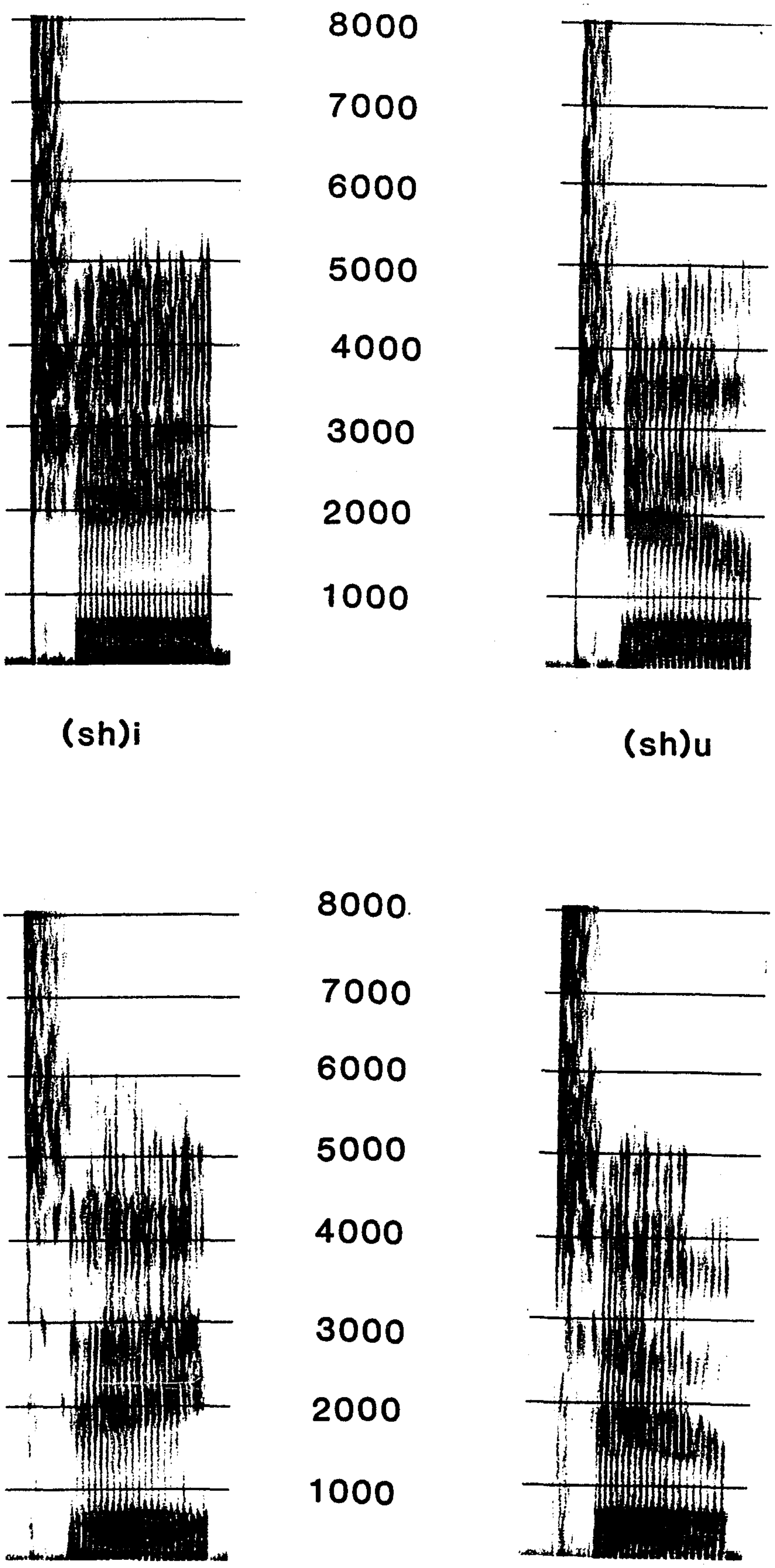
(s)i

(s)u

A7


8000
7000
6000
5000
4000
3000
2000
1000
(sh)i
(sh)u
8000
7000
6000
5000
4000
3000
2000
1000
(s)i
(s)u

A5



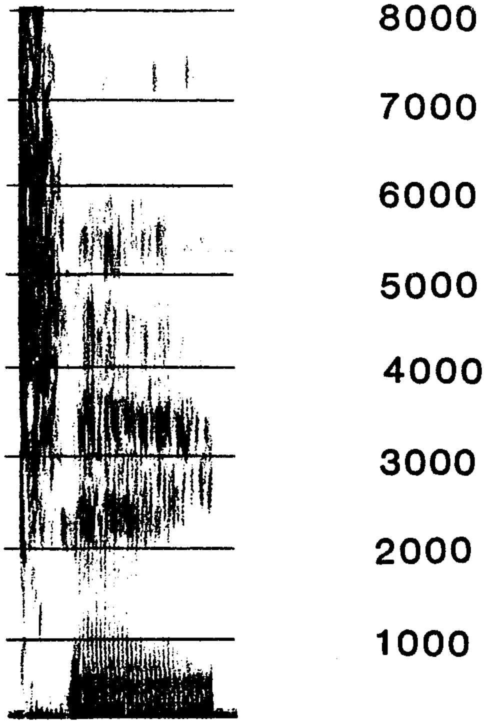
(sh)i

(sh)u

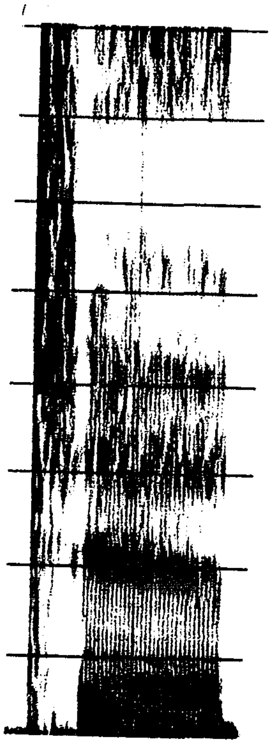
(s)i

(s)u

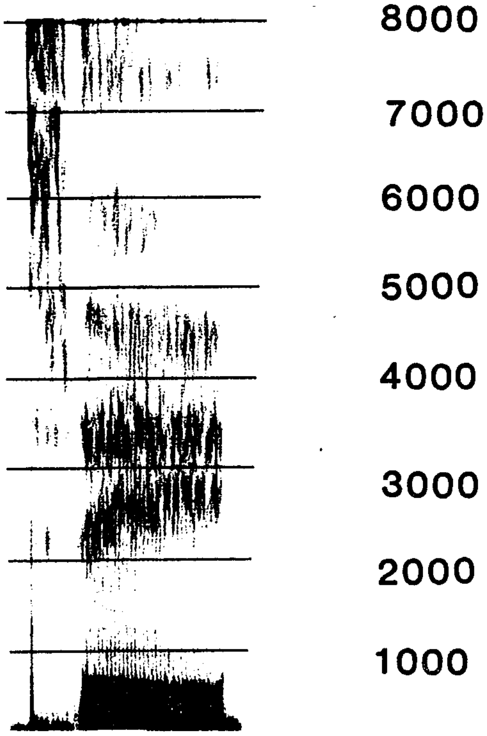
A2



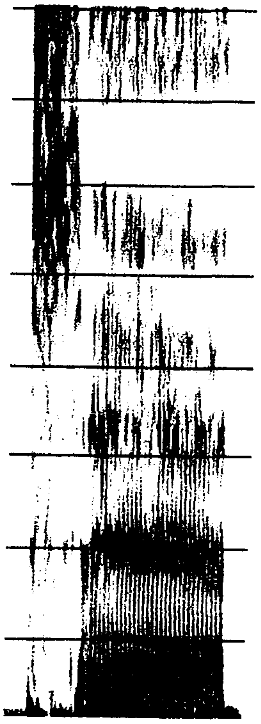
(sh)i



(sh)u

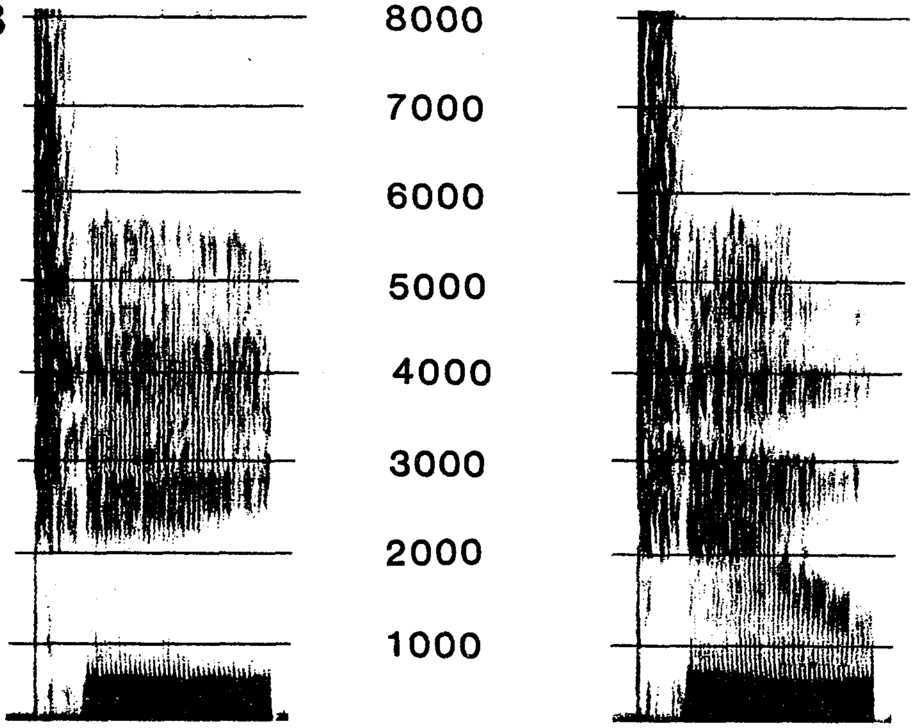


(s)i



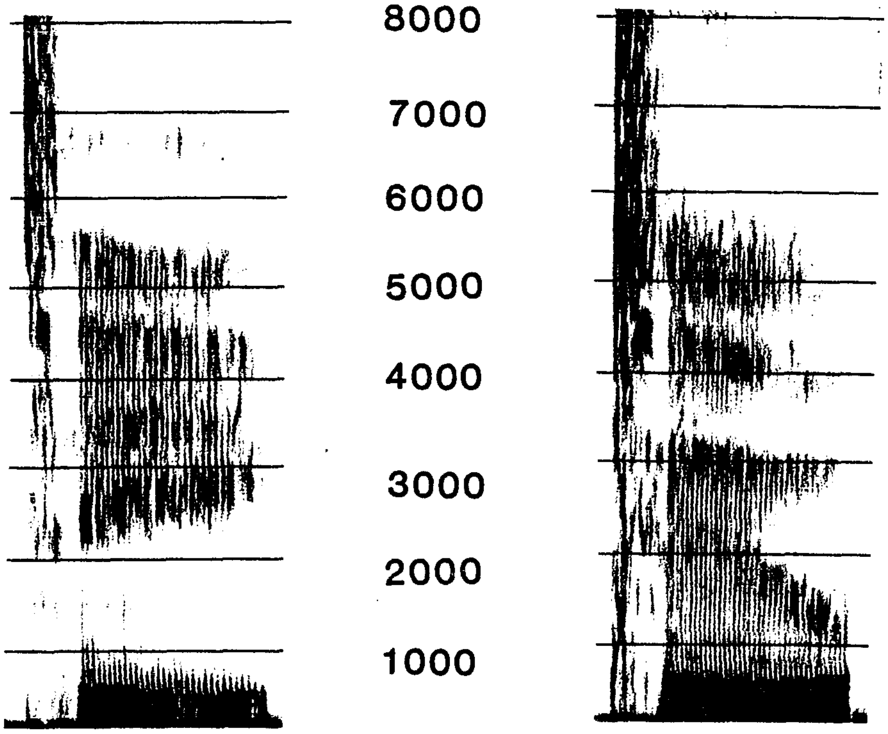
(s)u

A8



(sh)i

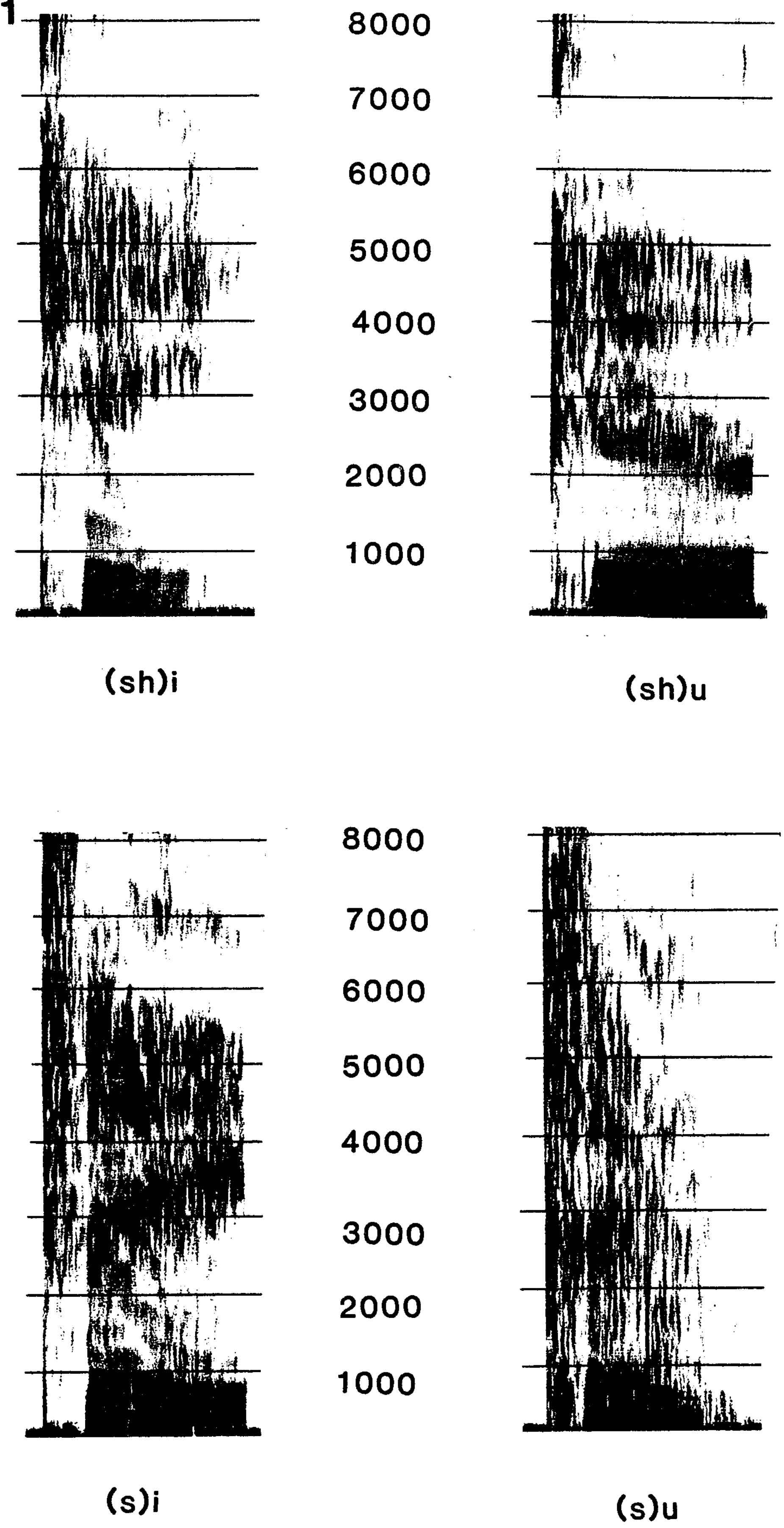
(sh)u



(s)i

(s)u

71



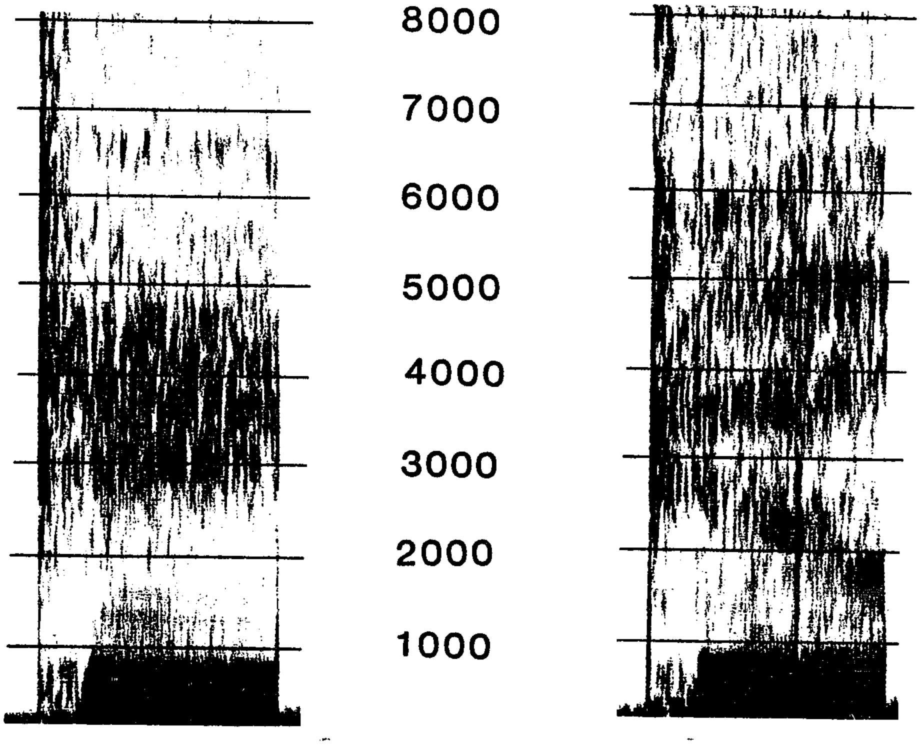
(sh)i

(sh)u

(s)i

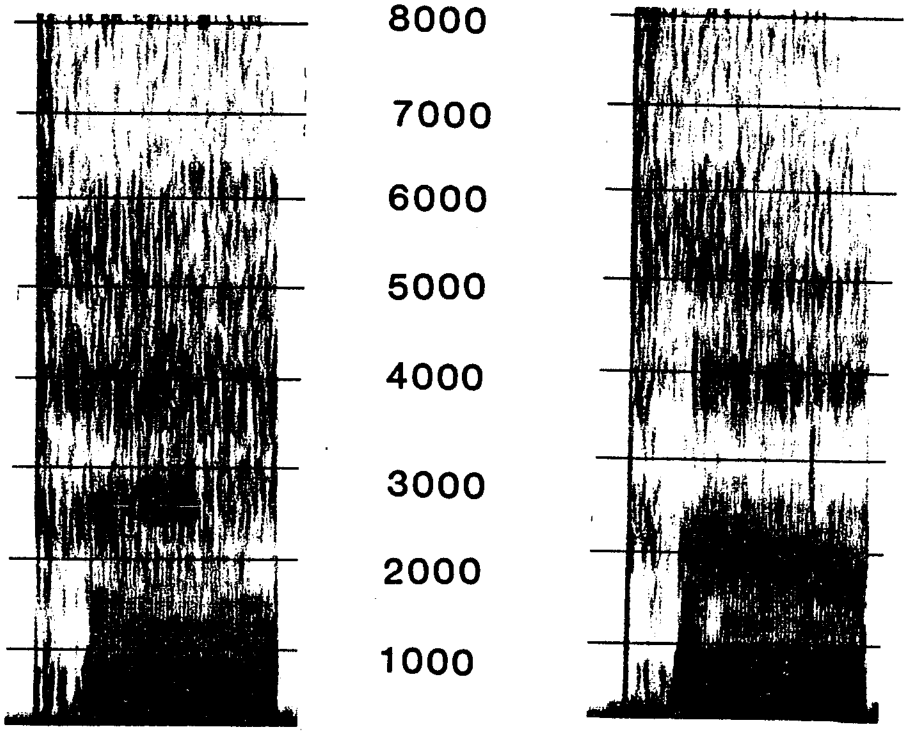
(s)u

72



(sh)i

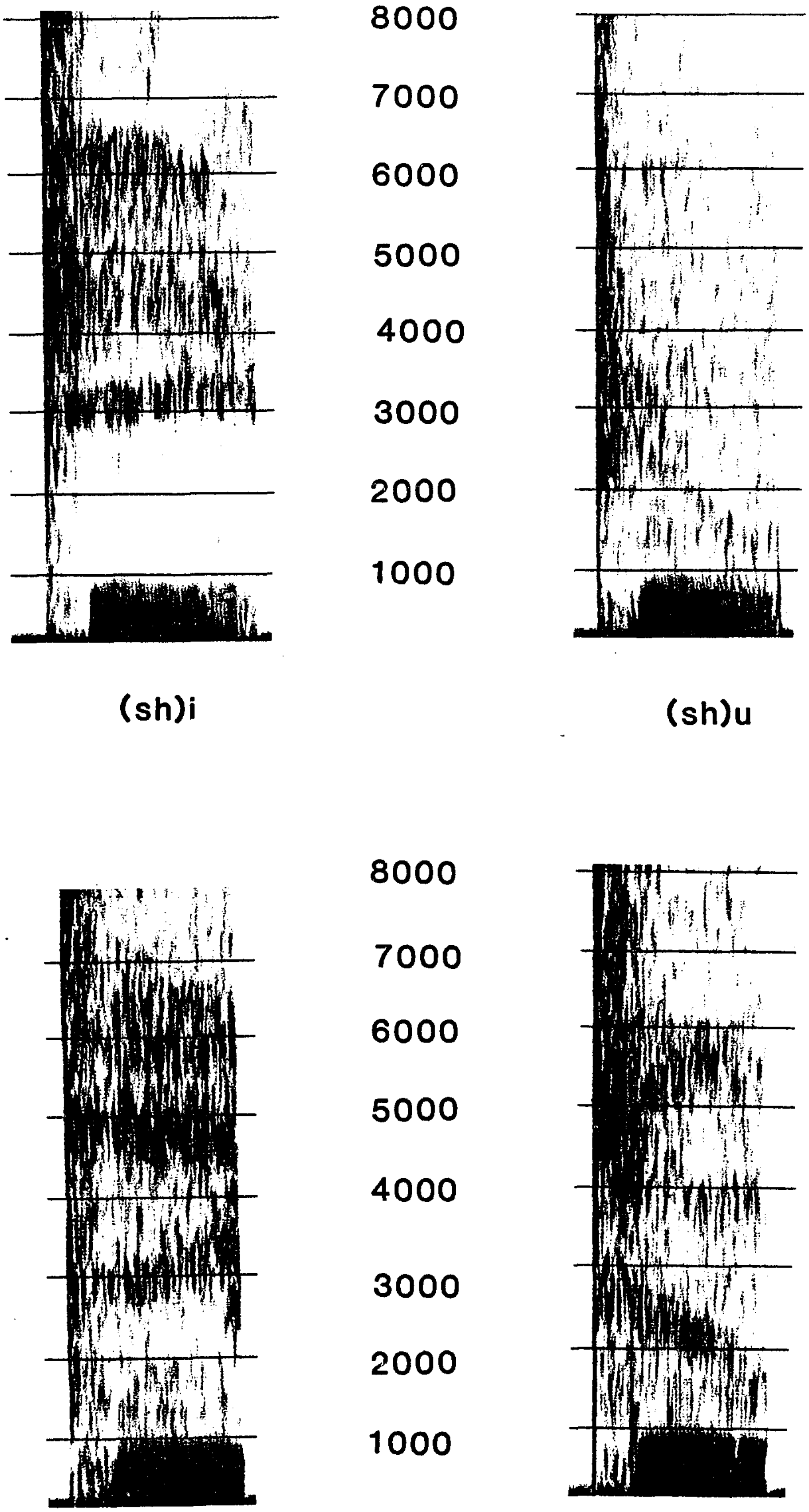
(sh)u



(s)i

(s)u

52



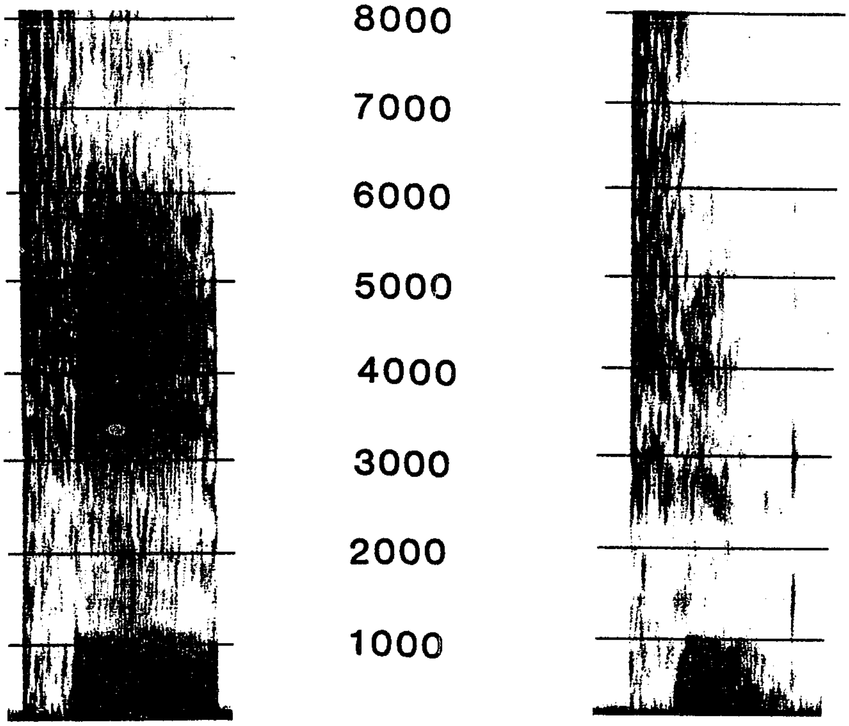
(sh)i

(sh)u

(s)i

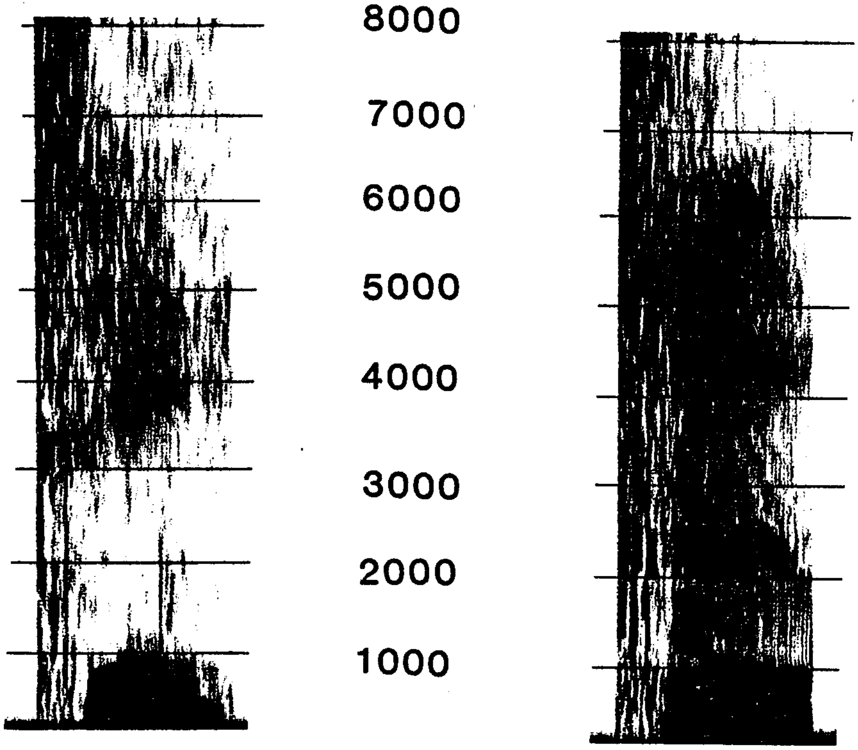
(s)u

56



(sh)i

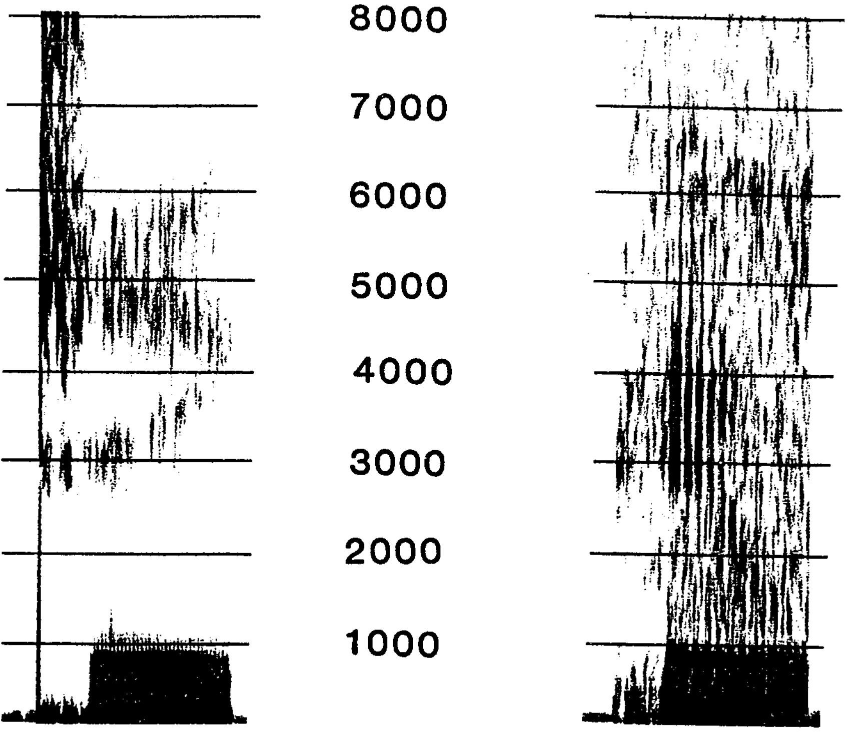
(sh)u



(s)i

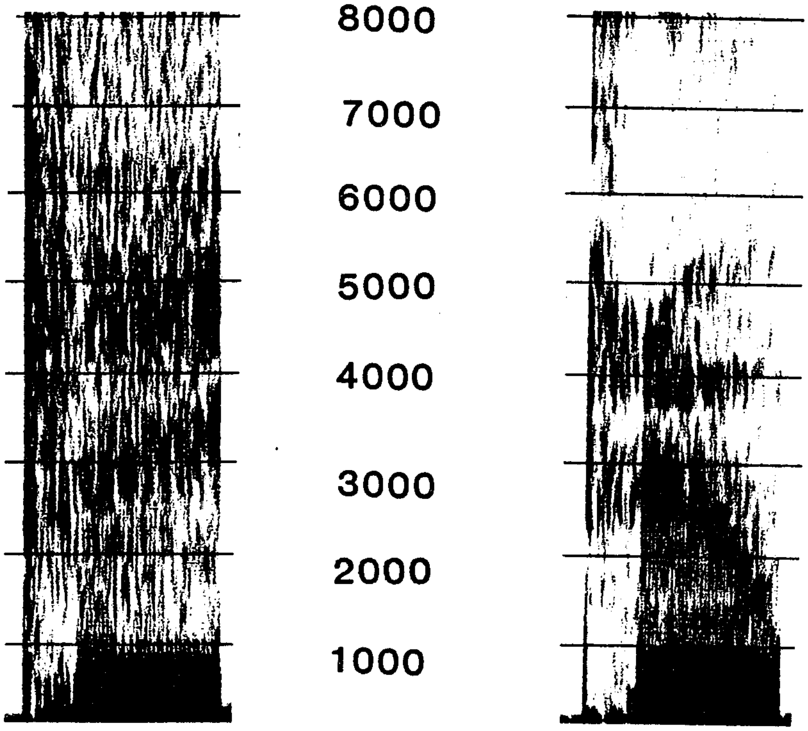
(s)u

44



(sh)i

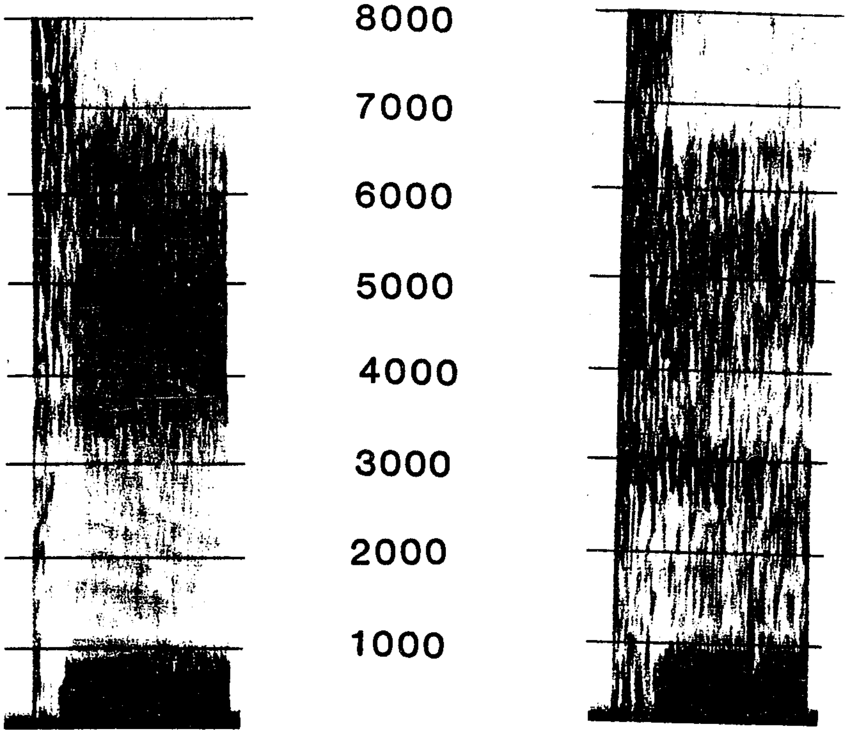
(sh)u



(s)i

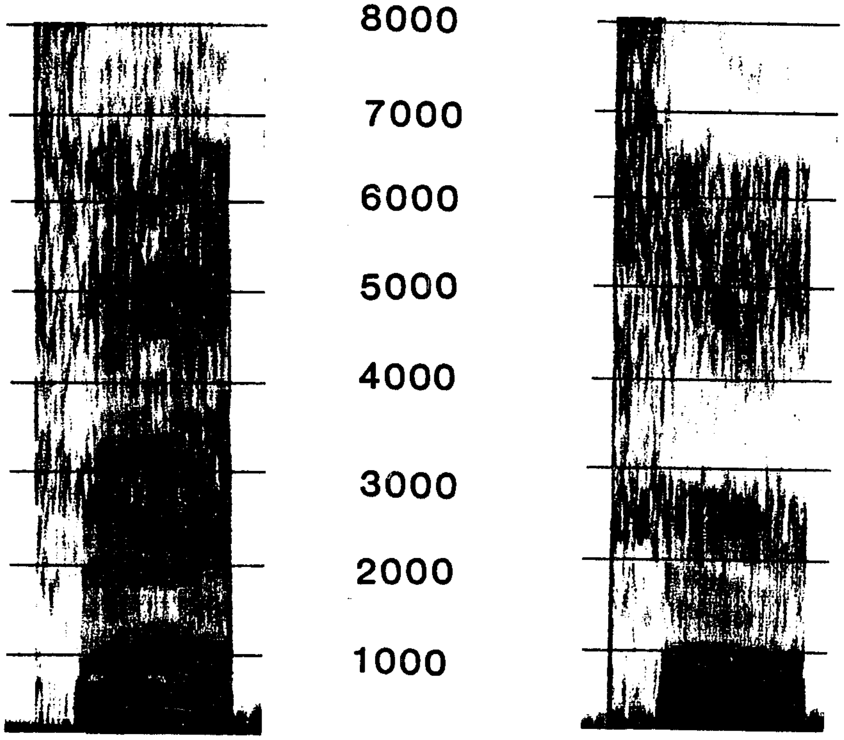
(s)u

41



(sh)i

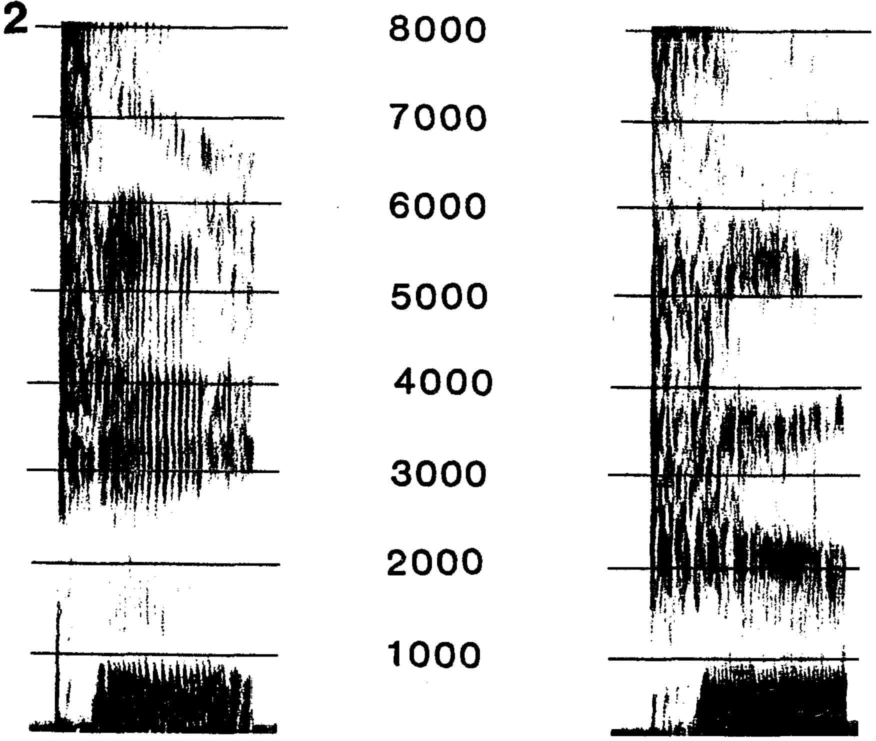
(sh)u



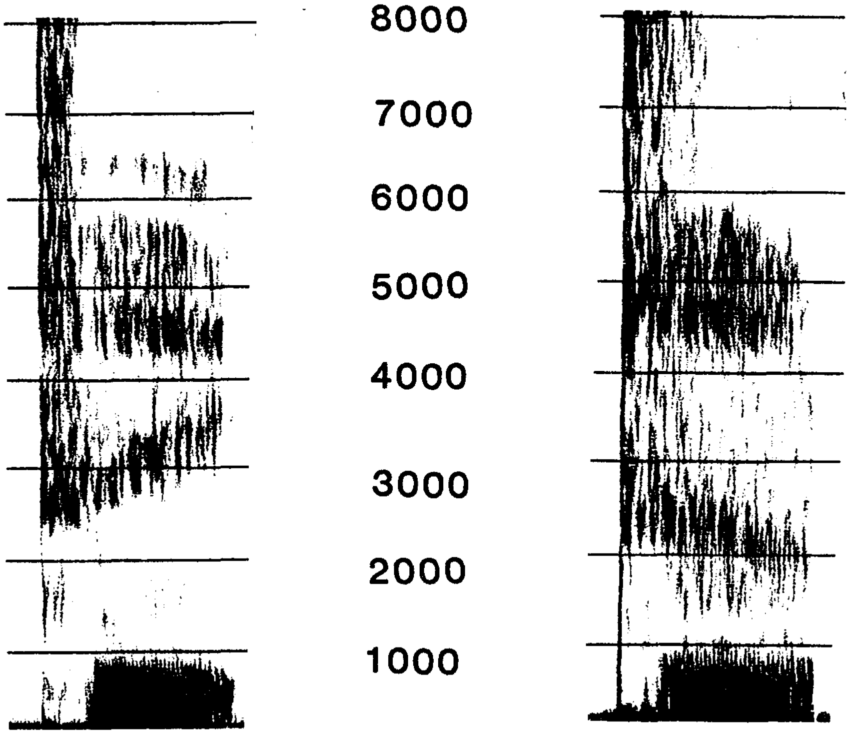
(s)i

(s)u

32

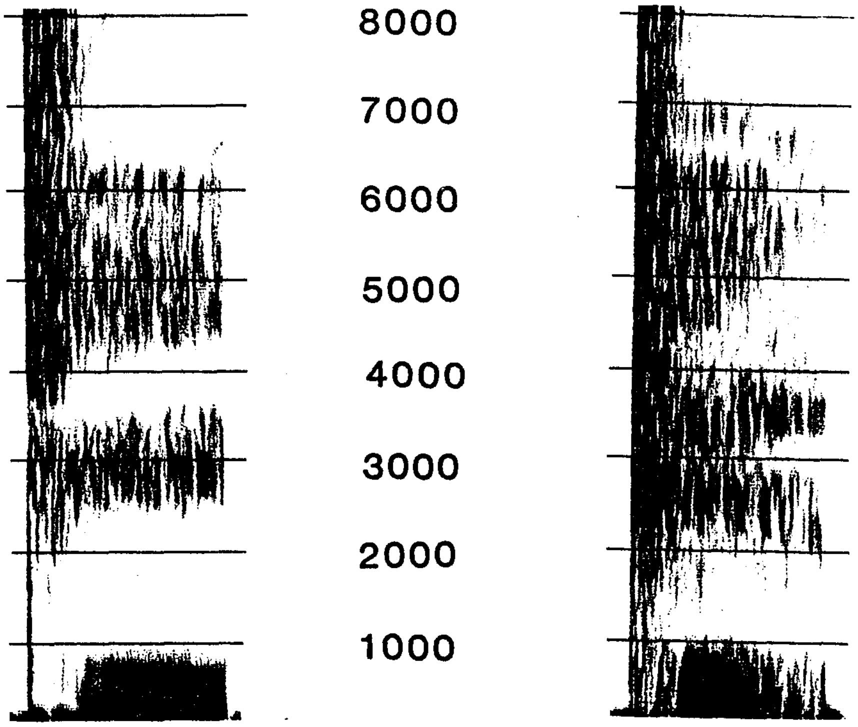


(sh)i (sh)u



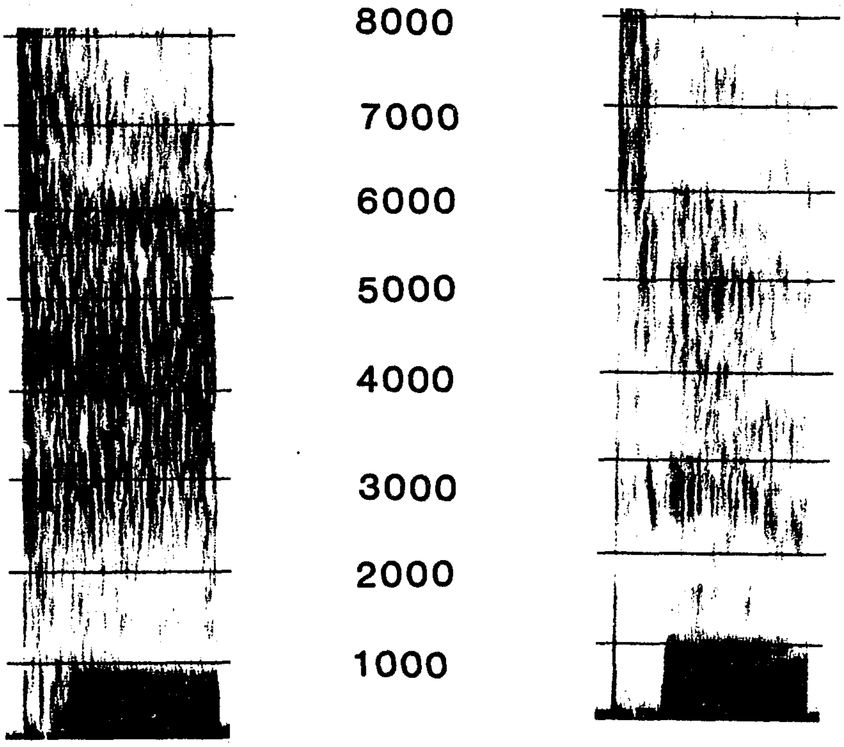
(s)i (s)u

35



(sh)i

(sh)u



(s)i

(s)u

References

Best, C.T., Morrongiello, B.A., & Robson, R.C. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception and Psychophysics, 28, 191-211.

Blumstein, S.E., & Stevens, K.N. (1981). Phonetic features and acoustic invariance in speech. Cognition, 10, 25-32.

Cooper, F.S., Liberman, A.M., & Borst, J.M. (1951) The interconversion of audible and visible patterns as a basis for research in the perception of speech. Proceedings of the National Academy of Science, 37, 318-325.

Cooper, F.S., Delattre, P.C., Liberman, A.M., Borst, J.M., & Gerstman, L.J. (1952). Some experiments on the perception of synthetic speech sounds. Journal of the Acoustical Society of America, 24, 597-606.

Darwin, C.J., & Bethell-Fox, C.E. (1977). Pitch continuity and speech source attribution. Journal of Experimental Psychology: Human Perception and Performance, 3, 665-672.

Dorman, M.F., Raphael, L.J., & Liberman, A.M. (1979). Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 65, 1518-1532.

Fant, G. (1960). Acoustic Theory of Speech Production. The Hague: Mouton & Company.

Fant, G. (1962). Descriptive analysis of the acoustic aspects of speech. Logos, 5, 3-17.

Finney, D.J. (1971). Probit Analysis. Cambridge: Cambridge University Press.

Fowler, C.A. (1980). Coarticulation and theories of extrinsic timing. Journal of Phonetics, 8, 113-133.

Fowler, C.A., Rubin, P., Remez, R.E., & Turvey, M.T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), Language production: Vol. 1. Speech and talk (pp.

373-420).New York: Academic Press.

Heinz, J.M., & Stevens, K.N. (1961). On the properties of voiceless fricative consonants. Journal of the Acoustical Society of America, 33, 589-596.

Hughes, G.W., & Halle, M. (1956). Spectral properties of fricative consonants. Journal of the Acoustical Society of America, 28, 303-310.

Joos, M. (1948). Acoustic phonetics. Language, 48(Suppl. 2), 1-136.

Kent, R.D. (1983). The segmental organization of speech. In P. MacNeilage (Ed.), Speech production (pp. 57-89). New York: Springer-Verlag.

Krause, S.E. (1982). Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults. Journal of the Acoustical Society of America, 71, 990-995.

Kunisaki, O., & Fujisaki, H. (1977). On the influence of context upon perception of voiceless fricative consonants. Annual Bulletin, 11, 85-91. (Tokyo: Research Institute for Logopedics and Phoniatrics.)

Levinson, S.E., & Liberman, M.Y. (1981, April). Speech recognition by computer. Scientific American, 244(4), pp. 64, 86.

Liberman A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Liberman, A.M., Delattre, P.C., & Cooper, F.S. (1952). The role of selected stimulus variables in the perception of the unvoiced stop consonants. American Journal of Psychology, 65, 497-516.

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

Mann, V.A., & Repp, B.H. (1980). Influence of vocalic context on perception of the /ʃ/-/s/ distinction. Perception and Psychophysics, 28, 213-228.

Mann, V.A., Sharlin, H.M., & Dorman, M. (1983). Children's perception of /s/ and /ʃ/: The relation between articulation and perceptual adjustment for coarticulatory effects. *Haskins Laboratories Status Report on Speech Research, SR-76*, 115-127.

Menn, L. (1978). Phonological units in beginning speech. In A. Bell & J.B. Hooper (Eds.), *Syllables and segments* (pp. 157-171). New York: North-Holland Publishing Company.

Menyuk, P., & Menn, L. (1979). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), *Language Acquistion* (pp. 49-70). Cambridge: Cambridge University Press.

Morrongiello, B.A., Robson, R.C., Best, C.T., & Clifton, R.K. (1984). Trading relations in the perception of speech by 5-year-old children. *Journal of Experimental Child Psychology*, *37*, 231-250.

Öhman, S.E.G. (1966). Coarticulated VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, *39*, 151-168.

Parnell, M.M., & Amerman, J.D. (1978). Maturational influences on perception of coarticulatory effects. *Journal of Speech and Hearing Research*, *21*, 682-701.

Pisoni, D.B. (1985). Speech perception: Some new directions in research and theory. *Journal of the Acoustical Society of America*, *78*, 381-388.

Simon, C., & Fourcin, A.J. (1978). Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America*, *63*, 925-935.

Soli, S.D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, *70*, 976-984.

Stevens, K.N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E.E. David & P.B. Denes (Eds.), *Human communication: A unified view* (pp. 51-66). New York: McGraw Hill.

Stevens, K.N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant & M.A.A. Tatham (Eds.), *Auditory analysis and the perception of speech* (pp. 303-330). New York: Academic Press.

Stevens, K.N., & Blumstein, S.E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *64*, 1358-1368.

Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, *3*, 32-49.

Whalen, D.H. (1981). Effects of vocalic transitions and vowel quality on the English /s/-/ʃ/ boundary. *Journal of the Acoustical Society of America*, *69*, 275-282.

Whalen, D.H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics*, *35*, 49-64.