INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Individual differences and the link between speech perception and speech production

by

Rochelle Suzanne Newman

August, 1997

A dissertation submitted to the Faculty of the Graduate School of the State University of New York at Buffalo in partial fulfillment of the requirements for the degree of Doctor of Philosophy

UMI Number: 9807324

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

There are many people to whom I owe a debt of gratitude. Thanks especially go to my committee chair and mentor, Dr. James Sawusch, who gave me the freedom and knowledge to follow my own interests. He has gone out of his way to provide advice and support, whether it was suggestions on setting up a lab, writing a review, or designing a course. He managed never to grumble at having to reboot the VAX after I had managed to lock it up yet again. In every way, he has been the model of a graduate advisor, and I hope I can come close to emulating him.

I also want to thank Dr. Joel Raynor, Dr. Lori Badura, and Dr. Gail Mauner, who with grace and good humor, served on my dissertation committee. Dr. LouAnn Gerkin for all her advice and help on job search. Special thanks go to Dr. Peter Jusczyk and Paul Luce, both for their quick response with regards to this manuscript and for all of the help and suggestions they have provided me throughout my graduate career. Their kindness and enthusiasm for working with a student from outside of their labs has given me three mentors, rather than just one, for which I am very grateful.

Thanks also to my family for their support, even when they did not quite understand what it was I was doing, and to my fellow graduate students and resident post-docs, with whom I have traded ideas, job information, and jokes. And most of all, to my husband, David Rossell, for helping me through the past few years, and for all the sacrifices he made along the way to help me get here.

TABLE OF CONTENTS

List of Tables

List of Figures

Abstract

A long-standing question in speech research concerns the degree of interrelation between speech perception and speech production. That is, are the representations used for these two different processes tightly linked, or possibly even identical? A related issue is whether there are reliable differences between individuals' perception which are related to the idiosyncrasies of their production.

Motor Theory (Liberman, Cooper, Harris & MacNeilage, 1962; Liberman & Mattingly, 1985) first proposed that speech perception takes place in reference to production. This would mean that the perceptual process makes use of the representations developed for production, and that differences between individual's productions should be reflected in their perception, as well.

A number of experiments have attempted to examine this issue over the years, but results have been quite variable. It is unclear whether this confusion is because the effect itself is variable, or whether more sophisticated experimental techniques might resolve the issue. The present set of experiments was designed to investigate this topic more closely.

The experiments reported here are modeled after an experiment by Miller and Volaitis (1989) in which they asked subjects to rate members

of a series for their category goodness. This allowed them to examine perceptual "prototypes" of a phoneme category for an individual listener. In the experiments described here, these perceptual prototypes were correlated with acoustic measurements of each listener's own productions . In the first experiment, listeners were asked to rate members of a VOT series ranging from /ba/ to /pa/ to /*pa/ (beyond a good "p"). Individuals who preferred tokens of /p/ with longer VOTs also produced longer VOTs in their own productions. Additional variance in the perceptual prototype was explained by production of /ba/. This suggests that voiced and voiceless stops provide separate, non-overlapping information about individual's mental representations, and that differences in perception are related to differences in production. A final finding from this experiment was that individual's perceptual prototypes tended to have more extreme VOT values than their own productions. That is, individuals seemed to demonstrate a "hyperarticulation" effect, as has been previously shown for vowels (Frieda, 1997; Johnson et al., 1993).

In Experiment 2, neither centroid of frication nor formant frequencies at onset of vocal pulsing demonstrated any correlation between perception and production in a /s/-/ʃ/ series. In the third experiment, a number of proposed cues were examined for stop consonants differing in place of articulation. Locus equations demonstrated no correlation between the two modalities for /b/, /d/, and /g/. Spectral moments and spectral peak differences showed no significant correlations on individual submeasures, but canonical correlations examining these entire sets of cues yielded high correlations. These canonical correlations were equal in size for the two sets of cues, suggesting that the sets are approximately equivalent descriptions of the information that listeners actually use.

The results from the set of experiments are not as clear as might be desired. The significant effect in Experiment 1 suggests that some links between perception and production do exist, and can be found with a suitable methodology. However, the variability across experiments suggests that this link is not especially strong, arguing against the notion that the modalities might share the same representations. Rather, it appears more likely that the link is indirect. Since the voice individuals have the most experience hearing tends to be their own, individuals' productions are likely to have a substantial influence on their perceptual prototypes.

CHAPTER 1

Speech Perception and Speech Production

A long-standing question in the area of speech research concerns the degree of interrelation between speech perception and speech production. Obviously, there are at least some connections between these two capabilities: For instance, human infants learn to speak their native language by hearing what other people produce. Thus, the infants must in some way associate the sounds they hear with the proper way of producing them, and this suggests some basic sort of linkage between the systems. But the controversy revolves around whether or not there are deeper connections than this, and whether or not it is likely that the same mechanism or representations might be used in both processes.

There are theories which have claimed explicitly that there is a common process that mediates both production and perception. For instance, motor theory (Liberman, Cooper, Harris & MacNeilage, 1962; Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985) argues that adults perceive speech by making reference to articulation. The earlier versions of the theory state specifically that listeners refer to how they themselves would articulate the sound in question. That is, perception takes place in reference to the individual's production. It follows, then, that there is a single source of information for both. Later versions of this theory have modified this approach. They instead argue that listeners perceive the intended gestures of the speaker through a rudimentary analysis-by-synthesis, and that this takes place in an innate speech-specific module. However, even this later version of motor theory does state explicitly that a common mechanism is involved in both production and perception: "[I]f speech perception and speech production share the same set of invariants, they must be intimately linked" (Liberman & Mattingly, 1985, p. 3). The authors even go so far as to claim that the word "link" really is not correct, since it implies that speech perception and production "though tightly bonded, are nevertheless distinct." Rather, they feel that "for language, perception and production are only different sides of the same coin" (Liberman & Mattingly, 1985, p. 30). This notion has been further supported by Ojemann and Mateer (1979; Ojemann, 1983) who found a site in the brain where electrical stimulation altered sequential facial movements as well as phoneme identification abilities. They argue, "Thus, nonverbal orofacial movements and phoneme identification share the same portion of the language cortex" and suggest that the two processes form "a sequential motor-phoneme identification (SM-PI) system for language, the central mechanism suggested by the motor theory of speech perception." (pg. 1402)[1].

In addition to the issue regarding use of a common mechanism, there is also a question as to whether differences across individuals in production might be related to individual differences in perception. Some phonemic distinctions can be articulated in multiple ways, with slightly different muscle movements (for example, see Perkell & Nelson, 1985; Perkell & Matthies, 1992; Johnson, Ladefoged & Lindau, 1993; Ladefoged, 1982, pg. 78). Different people may articulate the same sound with different combinations of muscle and articulatory action, and this might also influence what they expect to hear from other speakers. This notion would be expected from the standpoint of older versions of motor theory, since it claimed that the listener refers to his or her own articulation, rather than to some generalized notion of articulation.

Fowler's direct perception theory (Fowler, 1986) also suggests that perception and production may share a common mechanism. She suggests that listeners directly perceive the gestures of the speaker. This obviously

---

[1] However, many researchers disagree with Ojemann's claims in this regard. Brown (1983) Churchland (1983) and Studdert-Kennedy (1983) all argue that Ojemann's results do not require Motor Theory, and perhaps do not even support it. More specifically, even a perception/production connection could be because of either a motoric perceptual representation, or the reverse, a perceptual representation that is used for production (Frazier, 1983; Brown, 1983). Kent notes that studies of individuals with functional impairments do not bear out Ojemann's claim of a combined motor/phoneme identification area (Kent, 1983). Also, Cooper (1983) and Frazier (1983) argue that Ojemann's results may have been caused by stimulating a shared transmission line, rather than a shared processing site, which would not provide support for Motor Theory.

suggests some link between perception and production, but it might not be related to individual differences. Fowler seems to mean that listeners are perceiving the gesture the speaker made, not perceiving the gesture they themselves would have made. Only the latter would depend on the notion that individual differences in speaking would be related to perceptual differences. However, to the extent that perceiving the speaker's gesture requires learning, the individual's own articulations are likely to be a major factor in that learning (since they are what will be heard most often). Thus, individual differences could easily be incorporated into the notion that listeners perceive the speaker's gestures. But not finding these differences would not pose major difficulties with the theory.

In fact, most theories are fairly silent on this issue. While motor theory and direct perception would both be supported by finding individual perception-production links, only motor theory would have difficulties with a lack of this finding. And in fact, even these difficulties are not severe. Some research has already demonstrated that individuals with production difficulties can demonstrate normal perception (Aungst & Frick, 1964; MacNeilage, Rootes & Chase, 1967; Woolf & Pilberg, 1971; Haggard, Corrigall & Legg, 1971; Weiner & Falk, 1972; Waldman, Singh & Hayden, 1978; Strange & Broen, 1981; Broen, Strange, Doyle & Heller, 1983; Hoit-Dalgaard, Murray & Kopp, 1983; Rvachew & Jamieson, 1989;

however, see Travis & Rasmus, 1931; Kronvall & Diehl, 1954; Cohen & Diehl, 1963; Prins, 1963; Sherman & Geith, 1967; Weiner, 1967; Stitt & Huntington, 1969; Monnin & Huntington, 1974 for opposing results), and motor theory has been revised to take these findings into account. The most current instantiation of the theory claims that the perception-production link is innate, not learned. That is, whether a person actually has the opportunity to produce speech, he or she still compares incoming perceptual information to innate knowledge about speech production. As this innate knowledge is used in both production and perception, any differences in production should be present also in perception. However, although we know that individuals differ in their productions, we do not know anything about this innate knowledge. The theory is somewhat vague on this point, making it unclear whether such differences should exist at all. If this innate knowledge differs among individuals, then differences in production across individuals should correlate with differences in perception. But it is also possible that there are no differences in this innate knowledge, and any variation among individuals in production is due not to differences in the intended productions, but only to differences in performance (much like the performance difficulties of individuals with speech impairments). If this were the case, then the differences in production would not be expected to relate to perceptual differences. This provides a possible explanation for the absence of perception-production links, should future research fail to find them. Thus, while finding a correlation between differences in perception and production would support the theory, not finding such a correlation would not be a death-blow to the theory. It would, however, require that the theory make additional explicit claims regarding the source of production differences among individuals.

There are some additional theories which might fit in nicely with the presence of a perception-production link, but which do not depend on such a claim. Nearey's double weak theory (1992) argues that the perceptual system has knowledge about relations between speech-production capabilities and the resulting acoustic output (and that the targets the production system aims to produce are constrained by the kinds of things the limited perceptual system can readily decode). This requires some sort of perception-production link, but it does not depend on it being related to individual differences. Furthermore, Nearey suggests that listeners' representations of phonemes are abstract, and are not related in any simple manner to either the acoustic signal or articulatory gestures. Thus, the relation between acoustics and articulation is necessarily indirect. Finding that individual differences in production are related to those in perception might even be viewed as too strong of a relationship to easily mesh with such a theory, although Nearey has not discussed this issue explicitly. As with Fowler's theory, to the extent this knowledge about relations between production and perception comes about via learning, an individual's own productions might have a particularly strong influence. Such an argument could probably be used to incorporate the finding of such a link into the theory.

TRACE (McClelland & Elman, 1986; Elman & McClelland, 1986) is perhaps the only theory for which such a finding might be problematic. TRACE is an example of a connectionist model (that is, one which uses the interactive activation framework; see Rumelhart & McClelland, 1986; McClelland & Rumelhart, 1986 for an in-depth discussion of these models). In these models, information processing consists of the excitation and inhibition of large numbers of simple processing devices, or nodes. These nodes, and the links between them, make up a network that was originally conceived to be similar to the neural architecture of the brain. TRACE consists of three levels of nodes: the feature, phoneme, and word levels. That is, there are nodes that represent each of the possible phonemes in English, the features that make up these phonemes, and the words that are, in turn, made up of the phonemes. When perceptual information enters the model, it excites those nodes that are related to the input. An input will first excite the nodes representing those features present in the signal.

These feature nodes will then excite the phoneme nodes with which they are compatible. These phoneme nodes will excite the words that contain them, and will simultaneously inhibit other phoneme nodes. Thus, if an input consisted of features compatible with the phoneme /b/, the feature nodes would spread activation to that node, which would in turn inhibit other phoneme nodes (such as “p”) and excite relevant word nodes (such as “bag” and “bike”).

Because TRACE is based on abstract linguistic features (such as “acuteness” and “vocalicness”), it does not have individual differences built in. Differential experience could alter the weights on different features quite easily, however. Still, differences should only exist in the form of weighting changes, not in the form of differential features. Furthermore, TRACE is purely a perceptual model. It does not have any connections to production systems, nor does it have any obvious places where such a link could be added. In some sense, TRACE is no different than the myriad of other models which are silent on the issue of perception-production links. But unlike most models, which consist of modular components that can be altered without influencing other aspects of the model, the interactive nature of TRACE makes these additions quite difficult. Any change, however slight, alters the entire model. Thus, unlike many models, to which a perception-production link could be added without much difficulty, adding such a link to TRACE could greatly change the nature of the model itself. If there are these perception-production links, TRACE might require massive revisions to model the data.

The existence of individual differences

The notion that perception-production links can be examined through individual differences depends on the notion that individual differences actually exist. As mentioned above, there has been some research suggesting that different speakers do use different methods of articulating the same sounds (Perkell & Nelson, 1985; Perkell & Matthies, 1992; Johnson et al., 1993; Bell-Berti, Raphael, Pisoni & Sawusch, 1979; Ladefoged, 1982). A well-known example of this is the sound /s/, which can be produced with the tongue tip touching either the top of the mouth or the bottom row of teeth. There has also been a long history of work suggesting that individuals differ in their perception of speech. For example, Hazan and Rosen (1991) found a great deal of variability across different subjects' perception of synthetic speech series, especially for more complex, natural-sounding stimuli. Given this variability in both perception and production, it seems reasonable to examine whether the source of this variability might be the same in both cases.

Evidence from work with clinical populations and children

Most of the research that has looked for a link between perception and production has not been on normal speakers. The past fifty years have shown a wealth of studies examining whether children with articulation disorders also have difficulties in auditory discrimination tasks. While most of these studies make no claims about causality, there is an assumption that any child who has difficulty discriminating different sounds is unlikely to be able to produce these sounds correctly. Thus, finding a link between perception and production in these studies may not be too surprising. While it is important from a clinical standpoint (a misarticulating child with underlying perceptual problems will probably not be helped by pronunciation drills in the same way that a child with normal perceptual skills would be), it may not be as important from a theoretical standpoint. The primary theoretical issue is whether the representations used during normal perception and production are the same, or at least closely linked. Nonetheless, since clinical work makes up the majority of research related to this topic, it is important to gain an understanding of the prior findings. To that end, this section discusses these clinical studies in some depth.

There were a number of early studies that tentatively suggest the presence of a relationship between articulation errors and auditory discrimination abilities in children (see Weiner, 1967 for a review). That is, children who produce large numbers of articulation errors seem to have poorer auditory discrimination abilities as well, which may not be too surprising. However, this connection appears to be negligible in children who produce few or no errors.

In one of the earliest such studies, Travis and Rasmus (1931) found that grade-school children with articulation disorders made more discrimination errors than did normal speakers. Furthermore, the children who had the most severe production disorders generally failed to discriminate perceptually the same sounds that were the most difficult for them in articulation. Twenty years later, Kronvall and Diehl (1954; Cohen & Diehl, 1963) replicated these findings.

This prompted a wave of similar studies throughout the next decade. Stitt and Huntington (1969) was one of the few studies using adults, rather than children, and they found the same general results. They presented listeners with a wide variety of different tasks, and found that articulation ability correlated highly with speech discrimination, auditory identification and memory abilities in nearly all cases.

Sherman and Geith (1967) gave 529 children (all of whom had normal IQs and hearing scores) a speech sound discrimination task. They then gave an articulation test to the 18 children with the highest and lowest discrimination scores, on the assumption that any articulation difference between the groups would necessarily be correlated with their discrimination ability. They did find a significant difference between the groups, but unfortunately the groups also differed significantly in IQ scores, leaving open the possibility that the difference between groups on articulation ability might be an artifact of the testing situation, rather than an effect of discrimination ability differences.

Mange (1960) compared a group of normal children with a matched group of children who had difficulty articulating /r/ (but not /s/). He found that the two groups differed in their auditory pitch performance, but that this performance level was not correlated to the articulation scores within the groups. Conversely, the scores on a word synthesis test (an odd task involving the perception of three-phoneme words created by splicing together recordings of the individual phonemes in different environments) correlated with the degree of /r/ misarticulation, but did not differ between the two groups. Mange claimed that the pitch discrimination task was "related to normalcy or defectiveness of articulation but not to number of articulation errors. Synthesis ability appeared to be related to number of errors but not to normalcy or defectiveness" (p. 72). However, it certainly seems odd that a factor that correlated with number of errors would not also show a significant difference between a group that should have made multiple errors and a group that should have made very few.

Other findings were even less clear. Prins (1963) found that out of 22 possible correlations between speech discrimination and articulation, only 3 were significant at a .05 level (uncorrected for the number of tests performed). Furthermore, the correlation between the total number of articulation errors and the sound discrimination scores was not one of the ones that was significant. The following year, Aungst and Frick (1964) found that subjects who did not produce sounds correctly often failed to notice their mistakes, although they still performed normally on general tests of speech discrimination. They concluded that this suggests a link between children's speech production and their self-monitoring ability, but that this does not seem to have implications for perception in general. Lapko and Bankson (1975) came to the same conclusions following a similar study, but Woolf and Pilberg (1971) found no such correlation between production and the ability to evaluate or compare productions.

So, to summarize the results to date, several early studies suggested that articulation ability and discrimination ability may be linked. However, an approximately equal number of studies led to more ambiguous results. This negative trend became even stronger during the 1970s. Haggard, Corrigall, and Legg (1971) examined children who had difficulty articulating /s/, /r/, or both phonemes, but did not find that the children had difficulty perceiving the same sounds they had difficulty producing. The children with /s/ production difficulties did do worse discriminating the /s/ items, but they also had a tendency to do worse on the /r/ items, as well, suggesting overall poorer discrimination performance rather than a specific perception-production link. The authors conclude that whether an individual produces a sound correctly or incorrectly does not seem to correlate with their perception of that sound. But, how a person speaks (individual variation within the range of correct items) still may.

Weiner and Falk (1972) found no difference between misarticulating children and normal children's same/different discrimination of CV minimal pairs, either overall, or on the specific items the children had difficulty articulating. On the other hand, Marquardt and Saxman (1972), the same year, did find that misarticulating children made more discrimination errors than matched normals, although this may have been a more general testing problem, since these children also did poorer on a more general language comprehension task.

In contrast to these studies, Monnin and Huntington (1974) found evidence in favor of a perception-production link. They suggested that since the speech signal is normally redundant, removing this redundancy (by distorting the signal) might disproportionately increase the number of errors for children who misarticulate than for normal children, since normal children presumably would be more able to switch cues. Indeed, with mild to moderate distortion, the authors found that children who misarticulated the specific phoneme being tested tended to do worse on that discrimination, but did not do any worse on items which they produced correctly. (With large distortions, all three groups of children made a large number of errors.) The authors conclude that misarticulating children do have difficulty discriminating the sounds they themselves misproduce, but do not have a general perceptual problem.

Lewis (1977) examined the link between perception and production of particular linguistic features (specifically, those put forth by Halle (1964): +/- grave, diffuse, strident, nasal, voiced, and continuant). He compared groups of children on both naming and discrimination tasks and found that children with poor articulation had poorer discrimination scores overall, but the particular featural errors made in one task were not predictive of those in the other task. Waldman, Singh and Hayden (1978) also examined featural errors, but not only found no correlation between the number of featural errors the children made in each of the two tasks, but also that children with many articulation errors performed no worse perceptually than children with few.

Some of this variability in the literature may be because children who misarticulate are not necessarily a homogenous group. Strange and Broen (1981) tested 21 normal 3-year-old children on production and perception of /r/-/l/, /w/-/r/ and /w/-/b/ (control) contrasts. Although none of these children were labeled as misarticulators *per se*, /r/ is a difficult contrast to learn to produce, and many children at this age have difficulty producing it.[2] In comparison, /w/ is usually mastered by age 3 to 4, so the children were expected to have far less difficulty producing the /w/ phoneme. The children with the most difficulty perceiving /r/ tended to also have difficulty producing /r/, but the reverse did not always hold. Some poor producers did as well on the perception task as did the children who were perfect on the articulation task, and even those who made discrimination errors tended to have errors on all three contrasts rather than on just the contrasts involving /r/.[3] In other words, children differ: Some children have difficulty perceiving the distinction (and thus difficulty producing it), and some have production difficulty that is not correlated with perception problems. The authors also examined identification of an interpolated synthetic series, and found that the poor producers were less consistent in their responses to the /r/-/l/, and possibly /w/-/r/ series. It is still unclear, however, whether this variability is simply a sign of poor

---

[2] In fact, the authors report previous research by Sanders (1972) showing that this phoneme is not produced correctly by 90% of children until age 6, suggesting that many of the three-year-olds tested here are likely to have trouble with this contrast.

[3] On the other hand, they did tend to make *more* errors on the /r/ discrimination task. This could be taken to indicate the presence of a perception-production link, but could also simply mean that the /w/ contrast was less demanding in general.

attentional or test-taking abilities, or is actually a limit on these children's perceptual ability.

As a follow-up to this study, Broen, Strange, Doyle and Heller (1983) tested both normal and articulation-delayed 3-year-olds on minimal pairs consisting of the words wake, rake, lake, and bake (the pairs containing bake were considered control trials). The articulation-delayed group had more variable perception (which made it impossible to perform statistical tests to see if their mean values also differed). Furthermore, those subjects (in both groups) who neutralized the /w/-/l/ distinction in production had more variable perception of the distinction, and those articulation-delayed children who neutralized the /r/-/l/ distinction were likewise more variable in their perception of that distinction. The authors state that ". . . difficulty in the perception of a contrast may accompany production problems encountered by some but not all 3-year-old children" (p. 607). As in their first article, the authors claim that the relationship between perception and production exists, but is asymmetric.

Rvachew and Jamieson (1989) also found more variable perceptual performance for articulation-disordered children on fricative /s/-/ʃ/ ("seat" - "sheet") and /s/-/θ/ ("sick" - "thick") series. They found that adults showed a steeper slope, and more reliable identification than normal children, who in turn were more reliable than were articulation-disordered children. Like Strange and Broen (Strange & Broen, 1981; Broen et al., 1983), they concluded that some articulation-disordered children also have a perceptual disorder, but that some do not, and that these perceptual difficulties are specific to the misarticulated sound, rather than being general. However, as the authors did not try and relate perceptual performance to the actual pattern of misarticulations within each child, this latter conclusion is not backed by reliable evidence.

In another study focusing on the /r/ distinction, Hoffman, Stager, and Daniloff (1983) compared 12 children who consistently misarticulated [r] with 5 children who did not. All children were asked to repeat back sentences containing /r/-/w/ minimal pairs, and to identify all of the children's sentences (including their own) by a picture-pointing task. The misarticulating children did not perform any differently on the perceptual task for correct articulations than did normally articulating children, arguing against a perception-production link. Nor did they identify their own error productions better than other children's errors, going against the notion that the children were marking the distinction in a nonstandard manner. (Presumably, if the children were using a nonnormal cue to mark the distinction, they would have been able to use that knowledge to correctly perceive their own productions, just as they should have been unable to recognize the fact that they had made a mistake (as was discussed earlier).) However, data from individual subjects suggested that a subgroup of the children may have been marking the distinctions in an atypical manner, again suggesting that functional misarticulators may not be a homogenous group.

Hoffman, Daniloff, Bengoa and Schuckers (1985) followed up on this by examining children who maintained their productive impairment for [r] beyond the developmental period. All of the children could correctly identify “ray” vs. “way” when spoken by the experimenter, and the authors trained them to correctly identify the endpoints of a 7-item synthetic /r/-/w/ series. The children were then tested on their identification of the full series, and on their discrimination of pairs of stimuli (one pair contained two tokens of the same item, the other contained items 3 steps apart along the series). The misarticulating children took longer to learn the endpoints in the synthetic series than did normal children, and showed poorer performance on the series as a whole. The authors concluded that misarticulating children have poorer identification/discrimination of synthetic stimuli than normals on the sounds they have trouble producing, and that this may be because they use cues which are present in natural speech but not present in synthetic speech. That is, these children have latched onto different cues than do normal children. This could explain their articulation difficulties as well, as they would be producing phonemes according to the cues they had learned were important, rather than the cues viewed as important by society in general. Arguing against this, however, are their previous results (Hoffman et al., 1983) showing that most children did not seem to be making a non-standard contrast in this manner.

A few studies have attempted to train misarticulating children in perceptual tasks. Jamieson and Rvachew (1992) followed up their earlier results by training four misarticulating children (who demonstrated perceptual difficulties) on a perception contrast. Three of them managed to learn the perceptual distinction, and also showed concomitant production improvements, while the remaining child did not learn the series and failed to show any production improvement. Since successful training in the perception task aided these children's production abilities, it might suggest the existence of some sort of link. More recently, Griffiths and Johnson (1995) examined 2-year-olds' fricative productions in a similar manner. Although these children were developing normally, it is not uncommon to

---

have articulation that is not adult-like at this age.[4] The authors examined each child's productions, and then attempted to train the children perceptually on contrasts that they were still learning to produce (as well as on contrasts they had already mastered in production). They found that while children were able to learn other contrasts, all but one (out of eight) failed to learn the contrasts they had problems producing.[5] Some recent studies have examined low-level perceptual cues, rather than general identification performance. Hoffman, Daniloff, Alfonso and Schuckers (1984) compared VOT (voice onset time) values in perception with those from production for both normal and misarticulating children. They asked 12 kindergarten children (6 controls and 6 who were poor articulators) to repeat 9 sentences. Only one of the control subjects made as many as 3 phoneme errors on this, whereas the misarticulating children each made at least 6. However, none of the children in either group misarticulated the voicing of a prevocalic stop. There were 12 such prevocalic stops in the sentences (2 each of /p/, /t/ and /k/, 5 /b/ and 1 /g/), and the authors analyzed the VOTs of these items. They also created a 7-item synthetic /bi/-/pi/ series, and asked the children to point to the appropriate picture for each stimulus. The authors found that the misarticulating children

---

[4] According to Sander, 1972, the average 2-year-old does not produce any of the fricatives with consistent accuracy.

were more variable, and that there was a significant correlation between the perceptual category boundary and the production boundary (the half-way point between the mean voiced and voiceless productions) for these misarticulating children (r=.82, *t* =2.86, *p* <.05), although not for the control children (r=.11). This is highly suggestive of a link between perception and production. Yet, it is unclear why the correlation should be so high for misarticulating children and so low for normal children. Given that the normal children were not especially variable, six may not have been enough children for any correlation to appear. Perhaps a larger number of children would have shown a higher correlation.

Raaymakers and Crul (1988) found opposite results with an /s/-/ts/ series. Dutch children with articulation difficulties had poorer (and more variable) identification and an earlier phoneme boundary perceptually (that is, they require less silence to hear a /t/), but their successful productions had more silence to indicate presence of a /t/. This is directly opposite what one might expect if there were a link between perception and production. This effect was stronger in children who specifically had problems producing this distinction than in children with more general articulation difficulties, but was present in both. The longer silent periods

---

[5] Further work on this type of training procedure has been done with second-language learners, and will be discussed in the following section.

in production might not be too surprising, since these children are presumably less adept at the fine motor movements necessary to produce these sounds, and thus may produce them both more slowly and more variably[6]. But it is unclear why they would accept smaller silent periods than did the other groups as indicative of the presence of a /t/.

In another synthetic speech experiment, Lehman and Sharf (1989) tested adults and children of a range of ages on a /bit/-/bid/ ("beat" - "bead") series differing primarily on vowel duration (vowels are typically longer before /d/ than /t/ in English). Older subjects were less variable, had better discrimination scores, and had later perceptual boundaries. The authors also asked subjects to produce these items, and found that older subjects had a smaller separation between boundaries in production (that is, the difference between their average vowel durations for beat and bead were smaller). The only significant correlation between perception and production was in the variability. The authors suggest that these correlations may simply be missing the link, and that the tendency for category separation and variability in production and perception to decrease together with age suggests the presence of a link, regardless. But it may well be that younger children are simply poorer in their ability to

---

[6] There is a tendency to assume that this greater variability is simply an epiphenomenon of slower productions. However, work by Smith (1992) with normally developing children suggests that duration and variability may be separate indications of motor control.

perform the task, as they tend to be more variable in many other testing situations (including other speech production tasks; see Smith, 1992), and that the lack of correlations is really the important result.

Almost all of the studies above have examined misarticulating children. However, a few researchers have investigated different clinical groups. Hoit-Dalgaard and Murray (1983) examined 6 adult apraxic males on a b/p distinction. They found no apparent relationship between judgments of severity of the apraxia and the subject's VOT production data, or between the subjects' VOT boundaries in perception and their production. However, apraxia involves difficulty organizing purposive movements. Affected individuals often report that they know what they want to do but cannot organize the movements in order to do so correctly. Therefore, it is likely that these individuals' productions are affected not only by their representation of the item to be said, but also by their motoric difficulty. This suggests that these individuals' productions may not accurately reflect what they intended to produce, and thus it is not surprising that these productions would not be correlated with their perceptual representations. In fact, the apraxic participants often produced VOTs that were not even within the proper range for the phoneme.

Different findings have been shown for developmental apraxics, however. Groenen, Maasen, Crul and Thoonen (1996) presented speech continua varying in place-of-articulation to both apraxic and normal children. While the apraxic children in this experiment had similar identification functions to normal children, they had poorer discrimination functions, which the authors take to indicate normal phonetic processing paired with deficient auditory processing. (However, according to nearly all theories of speech perception, deficient auditory processing would be expected to result in phonetic errors, as well, making this distinction by the authors tenuous, at best.) Following this comparison across groups, the authors examined the types of errors individual apraxic children made, and correlated this with their discrimination scores. They found that children who made proportionally more errors involving place-of-articulation tended to demonstrate poorer perception for the place continuum, as well, suggesting a link between perception and production at the level of individual subjects.

MacNeilage, Rootes and Chase (1967) examined a patient with severely impaired somesthetic perception. In addition to insensitivity to pain, this individual had poor temporal and spatial resolution in muscle activity, leading to difficulties in swallowing, speaking, and other fine motor activities. Her speech production was fairly accurate for vowels and nasals (which may require less precise muscle movements), but extremely deficient for all other speech sounds. Yet despite these shortcomings, her speech perception seemed relatively preserved. The authors argue that reference to normal motor information does not appear to be a prerequisite for perception (although it may still play a role in normal subjects' perception).

There is one last paper that examines a clinical population, although not one involving individuals with production difficulties. Ojemann and Mateer (1979; Ojemann, 1983) examined 4 patients undergoing left temporal lobectomies for medically intractable epilepsy. They performed stimulation mapping, and found that nonverbal orofacial movements and phoneme identification share the same portion of the language cortex, suggesting that the two might be related functions. They suggest that this portion of the brain is responsible for both sequential motor movements and phoneme identification, and that it is "the central mechanism suggested by the motor theory of speech perception, which this association supports" (p. 1402). However, this finding has been questioned by a number of researchers. Some (Cooper, 1983; Frazier, 1983) argue that Ojemann may have stimulated a shared transmission line, rather than a shared processing site. Furthermore, even if there is a shared processing site, it could be because of either a motoric perceptual representation (as motor theory suggests), or a perceptual representation that is used for production, which would be inconsistent with such a theory (Frazier, 1983; Brown, 1983). In addition, studies of individuals functional impairments are not consistent with a combined motor/phoneme identification area (Kent, 1983).

To summarize this section, it appears that children who misarticulate a given sound may have difficulty discriminating between that sound and other, similar phonemes. Certainly, this seems to be the case for some children, if not for all. While it is impossible to make a statement of causality, it seems reasonable to suggest that a difficulty in perceiving a particular distinction might be the cause of the difficulty in producing it. After all, speech distinctions are specific to the language being learned. If a child does not perceive a distinction correctly when it is being produced by the adults around her, it is rather unlikely that she could nonetheless learn to produce it correctly herself (especially given that "correctly" in this context really means "in accordance with the societal norms").

What is unclear is the extent to which this necessitates the existence of a link between perception and production. Obviously, it is very difficult to learn to pronounce a sound correctly if you cannot hear it. But were that all that was meant by having a perception-production link, the issue would be rather uninteresting. What is really in contention is whether individuals who have normal production and perception still make reference to the same mental information regardless of whether they are speaking or talking. That is, whether the representations used during perception and production are the same, or at least closely linked. While the clinical research may suggest that they are, the connections found in this literature can easily be explained by assuming that some misarticulating children simply mishear. In fact, the results here need not have any implications for adult speakers at all.

Evidence from cross-linguistic work and work with second-language learners

Although not as extensive as the literature on articulation-disordered children, there is an important literature examining links between perception and production in both second-language learners and in non-English speakers.

Flege (1993) assessed the degree to which English learners from mainland China and Taiwan were able to use vowel duration as a cue to final stop voicing. In English, the duration of a preceding vowel varies with the voicing of the following stop, such that vowels are longer when followed by a voiced /d/ than when followed by a voiceless /t/. Chinese does not allow any final stop consonants, however, and Taiwanese only allows voiceless stops. Thus, the use of vowel duration as a cue to stop voicing should be a novel distinction to both groups of speakers. However, since Mandarin Chinese does not even allow for final stops, they may be less likely to pay attention to word-endings than the Taiwanese speakers, and thus less able to pick up the final t/d distinction. Flege tested the assumption that non-native speakers would show discontinuities in imitated vowel durations only if they covertly categorized word-final stops in the consonant-vowel-consonant stimulus as /t/ or /d/. That is, their productions would only show a categorical distinction in vowel duration to the extent that they were capable of perceiving this voicing distinction, suggesting that perception of non-native contrasts leads production. The data for groups of subjects differing in experience with English was consistent with this hypothesis, but data from individual subjects did not match this pattern.

Flege and Schmidt (1995; Schmidt & Flege, 1995) examined native Spanish speakers who learned English later in life. Spanish /p/ is produced with a short lag between release of pressure in the vocal tract and onset of vocal fold vibration, whereas English /p/ is produced with a much longer temporal lag. The authors examined both productions and perception of /p/ for these subjects at different speaking rates, and looked for correlations between them, as a way of determining the extent to which the subjects had successfully learned the new phonetic category. Out of 20 potential correlations between perceptual and production measures, only 2 were significant at the .05 level (uncorrected for the number of correlations). Both of these involved, as the perceptual measure, the effect of speaking rate on the lower limit of temporal lag considered acceptable by the subjects. That is, the authors did *not* find a correlation between absolute measures of VOT in production and perception, but instead found a correlation between the degree to which subjects were affected by speaking rate in perception and how they produced these items normally. It is unclear why this type of correlation would be significant when other, more obvious correlations failed the significance test. Given the large number of correlations performed, it is certainly possible that the significant effects were spurious.[7] One possible reason for the lack of a correlation in absolute measures of VOT comes from a related study by the same authors (Schmidt & Flege, 1996). They reported that production values for English monolinguals had little intersubject variability for initial /p/ productions. If there was little variability among their subjects, it would be very difficult to find a significant correlation across subjects.

Flege and Eefting (1986) found that children (in both English- and Spanish-learning environments) have significantly earlier perceptual boundaries on a /t/-/d/ VOT continuum than adults in their respective linguistic cultures. This same difference was found in production. That is,

---

[7] Although the correlations between perception and production in Native English speakers were not reported in these studies, Flege recently re-analyzed his data in this regard (personal communication, 1996). His findings are highly consistent with the results reported in Experiment 1; that is, there was a moderately-high correlation (r=.5361) between these subjects' productions of /p/ and their preferred VOT for synthetic /pi/ syllables. However, this only held for the perception of slow-rate syllables, not for items at a fast-rate.

children tended to produce /t/ with shorter (more "d"-like) voice onset times than adults. (Although the production effect did not reach statistical significance in this study, it was significant in a similar study; Flege & Eefting, 1987). The authors note that, at least for English-speakers, perceptual boundaries tended to fall intermediate to speakers' productions. That is, whatever voice onset times an individual produced in their "t" and "d" tokens, their category boundary fell roughly in the middle. This led the authors to speculate that perhaps the reason adults require longer VOTs perceptually to hear an item as voiceless than do children is *because* they produce the stops with longer VOTs. As they point out, this would imply "a very close link between those aspects of a phonetic representation which specify motoric control and perceptual processing" (p. 165).

Another group of studies in the area of second-language learning has looked at Japanese speakers learning the English /r/-/l/ distinction. Yamada and Tohkura (1990) argued that perception and production are strongly related in these speakers, and more recent research has focused on trying to examine this relationship more closely. Bradlow, Pisoni, Akahane-Yamada and Tohkura (1997) trained Japanese speakers on the /r/-/l/ distinction with a perceptual identification task. They found that not only did this training improve the participants' perception, but it also resulted in improved production: Their /r/ and /l/ productions both sounded better, and were more intelligible, to native English speakers following the training. While each participant showed some improvement in both perception and production, the degree of improvement was quite variable. There was no correlation between the amount of learning in the two modalities. That is, subjects who showed a greater perceptual improvement did not necessarily show more production improvement as well. The authors state, "we observed a link between perception and production to the extent that perceptual learning generally transferred to improved production . . . [but] we found little correlation between degrees of learning in perception and production after training in perception, due to the wide range of individual variation in learning strategies. . . . Taken together these findings support the hypothesis that learning in perception and production are closely linked" (p. 2307). But while these findings are predicted by models such as those discussed above (Fowler's direct realist model, Liberman *et al.*'s motor theory), these models have no explanation for the lack of correlation between degrees of learning.

In conclusion, it appears that as second-language learners begin to distinguish non-native phoneme contrasts perceptually, they also begin to show differences in production. This might suggest that a similar representation is being used in the two processes. However, results from nonnative speakers tend to be extremely variable, making perception-production links difficult to find.

Evidence from work with adaptation

One further area of research is relevant to the issue of perception-production linkages. During the 1970s and early 1980s, a great deal of research involved the method of selective adaptation (Ades, 1974; Ades, 1977; Ainsworth, 1977; Diehl, 1981; Diehl, Kluender & Parker, 1985; Elman, 1979; Ganong, 1978; Garrison & Sawusch, 1986; Jamieson & Cheesman, 1986; Roberts & Summerfield, 1981; Samuel, 1986; Samuel, 1988; Samuel, 1989; Samuel, Kat & Tartter, 1984; Sawusch & Pisoni, 1978; Sawusch, 1976; Sawusch, 1977; Sawusch & Jusczyk, 1981; Simon & Studdert-Kennedy, 1978; Eimas & Corbit, 1973; Eimas, Cooper & Corbit, 1973). Selective adaptation involves repeatedly presenting a subject with a single auditory stimulus. This repeated presentation causes listeners to then perceive a new auditory signal differently. At first, selective adaptation was viewed as fatigue to a phoneme detector, similar to the aftereffects found following visual receptor fatigue. For example, after repeated presentation of the sound /ba/, the /b/ detector becomes fatigued, and responds less strongly to /b/ tokens. An item which had been ambiguous between /b/ and /d/ (that is, which caused both /b/ and /d/ detectors to fire) will now be perceived as a better example of /d/ (because only the /d/ detector would now be firing).

William Cooper (1974) examined whether adaptation to a perceptual stimulus influenced *production*. He presented subjects with repeated presentations of either /bi/, /pi/, or /i/ (a neutral adaptor) and examined the voice onset time (VOT) of subjects' productions of /pi/ and /bi/. The phonemes /bi/ and /pi/ differ primarily in the timing relationship between the release of the consonant and the onset of voicing. There is a greater latency (or VOT) in /pi/ and a shorter latency or VOT in /bi/. Following adaptation with /pi/, listeners' productions of /pi/ had a shorter latency than they did following adaptation with /i/. In other words, after hearing the syllable /pi/ repeatedly, listeners' productions of /pi/ were more "/bi/-like." However, there was no significant shift for /bi/ productions. Cooper argues that the mechanisms for these two consonants operate separately from one another, and that "the adaptation effect represents the fatiguing of a single mechanism utilized during both speech perception and speech production" (p. 231).

Cooper and Lauritsen (1974) extended these findings, by showing that adaptation with /pi/ also has effects on the production of /ti/, as has been found with perceptual adaptation. "The results for the [ti] utterances indicate that the stage of processing subserving both the perceptual and motor systems of speech" involves "processing information about the voicing property of the consonant" -- thus, at the level of processing involved in both perception and production, adaptation with /pi/ actually fatigues a detector for the abstract linguistic feature "voiceless", rather than a detector for /pi/ itself (p. 122).

In yet a further study, Cooper and Nager (1975) found that adaptation with [rəpʰi] has the same effect on productions of [rətʰi] as did productions of /pi/ on /ti/. However, Summerfield, Bailey and Erickson (1980) failed to replicate this result when using subjects' own productions as adaptors. Cooper, Ebert, and Cole (1976) likewise found no perceptuomotor adaptation on production of /sti/ following multiple presentations of that syllable, even though this did result in perceptual adaptation of a /si/-/sti/ continuum.

Cooper, Blumstein, and Nigro (1975) examined the possibility of the converse effect: That is, whether repeatedly *producing* a syllable would have effects on perception (even when the listeners were prevented from hearing their own productions by white-noise). Three out of four subjects who repeated the syllable /bæ/ showed a shift in their perceptual category boundary for a synthetic /bæ/-/dæ/ series. In addition, three out of eight subjects showed a large perceptual adaptation on this series after repeatedly whispering the sequence /bæ/-/mæ/-/væ/, although the shift did not reach significance across the group.[8] The three listeners who showed the effect had also shown relatively large perceptual adaptation, and thus the authors suggest that there is a perceptual-motor effect for some listeners, but "its appearance depends on a strongly adaptable speech processing system, present in only some of our subjects" (p. 95).

Cooper, Billings, and Cole (1976) investigated a larger number of series in the interest of extending these results. They examined /si/-/sti/, /ba/-/wa/, and /ba/-/pɑ/ distinctions, and found effects of whispering the 2 syllable sequence /sti/-/stu/ on a /sti/-/si/ continuum, but no effect of producing /si/-/su/ on the same perceptual continuum. On a /ba/-/wa/ continuum, they found adaptation following productions of /wa/-/ya/, but not following productions of /ba/-/da/ (whereas this sequence does produce perceptual selective adaptation effects). No effects of adaptation were obtained with a [ba]-[p^{h}a] continuum. The authors suggest that this voicing distinction may not be processed in the same manner for whispered speech (where all sounds are effectively voiceless) as for normal speech. Still, even accepting this explanation for the final series, the results overall were highly variable. The authors admit this, claiming ". . . these results

---

[8] Although speaking in noise should have prevented listeners from hearing their own speech by air conduction, it might not have prevented listeners from receiving some auditory information by way of bone conduction. Whispering, however, does not engage the vocal tract, and is thought to prevent the possibility of bone conduction.

provide some support for the existence of an auditory-motor processor which serves both speech production and perception. However, in comparison with the results of tests using a strictly perceptual adaptation paradigm, the articulatory effects on speech perception are fraught with asymmetries, inexplicable in terms of any known concepts of speech processing" (p. 231).

Shuster and Fox (1989; Shuster, 1990) examined the final possibility, motor-motor adaptation. Here, listeners repeatedly produced one speech syllable, and then produced a single token of a second syllable. The authors found consistent effects of adaptation, and argued that both this task and perceptuomotor adaptation tapped into the same mechanism, one used for both perception and production of speech.

Overall, there appears to be some tendency for adaptation in either perception or production to influence the other. However, this effect is somewhat variable, and may depend critically on the specific tokens or tasks involved, or on the specific individuals, weakening any possible conclusions. Furthermore, researchers have failed to replicate some of these results, again making conclusions based on these studies somewhat suspect.

Evidence from work with normal populations

In addition to the work with adaptation, and the work with nonnormal populations discussed above, there is some experimental evidence that supports the contention that how normal individuals produce a given contrast will be related to the way they perceive it. As this body of literature is more directly relevant to the issue at hand, it will be discussed in slightly more depth.

In the first such study, Bell-Berti, Raphael, Pisoni, and Sawusch (1979) examined EMG recordings of three speakers producing the phonemes /i, I, e, E/. Linguists generally refer to the difference between /i/ and /I/ (and between /e/ and /E/) as a difference between "tense" and "lax" vowels. Bell-Berti and colleagues found that there were two different ways of producing the tense-lax distinction, and that different speakers used different strategies. The authors then presented 137 listeners with an /i/-/I/ continuum, both in a straight labeling paradigm and in an anchoring paradigm. They found a bimodal pattern of results, with some subjects showing a much greater anchoring effect than others.[9] Finally, 10 subjects participated in both the EMG task and the perceptual task. Four of

---

[9] In an anchoring study, a single item (here, the /i/ endpoint) is presented more often during the course of the experiment than are other item. The result is that this item serves as a referent, and other members of the continuum are contrasted with it. The anchoring effect here, then, is that other members of the series seem less /i/-like (more /I/-like), and thus that the category boundary is shifted towards the /i/ anchor.

these subjects used a more traditional production strategy, and all showed large anchoring effects; the remaining six subjects used the alternative production strategy, and showed much smaller anchoring effects. That is, there appeared to be bimodal distributions in both the production and perception tasks, and a high degree of correlation between the speech production strategy used by each subject and their performance on an anchoring task.[10] The manner in which individual subjects produced a given contrast was highly correlated with those subjects' perceptual data.

These results provide strong evidence of some sort of perception-production link, and of the existence of individual differences in perception and production tasks. However, the authors only tested 10 subjects who participated in both conditions. Furthermore, the authors could not find any systematic differences in the acoustic measures of productions by the two groups of subjects. Nonetheless, these results certainly suggest that individual differences in production and perception may well be related.

While Bell-Berti *et al.* found a connection between a measure of articulation (EMG data) and one of perception, there have been other

---

[10] It is not clear how using one production strategy instead of another would make an individual less resistant to anchoring effects. The authors suggest that the anchoring effect only appeared in individuals for whom the vowel stimuli represented adjacent categories in phonetic space. That is, since the anchoring effect was larger for individuals who made the tense/lax distinction on the basis of tongue height than it was for those who made the distinction via tongue tension, the tense and lax vowels were members of adjacent categories for the former group, but not in the latter group. This would suggest that individuals differ not only in production strategies and perceptual prototypes, but also in the complete layout of their phonetic space. However, there has been no further evidence in support of this suggestion.

studies that have looked for a correlation between *acoustic* measures of production and perception. The first of these experiments was by Bailey and Haggard (1973). They gave 34 subjects a series of synthesized speech stimuli ranging from /kIl/-/gIl/ ("kill" to "gill"). The primary difference between /k/ and /g/ is in their voicing: /g/ is considered a voiced consonant, whereas /k/ is voiceless. Voice onset time (VOT) is generally considered to be the primary cue to this distinction. This cue will be described in more detail later. For now, the important point is that it is possible to make items that are intermediate between /k/ and /g/ on this cue, and thus to make a series ranging from /k/ to /g/. There were 10 stimuli overall, consisting of five different voice onset times (VOTs) and two different values of onset fundamental frequency. (The fundamental frequency, or F0, changed over the beginning portion of the syllable and reached the same steady-state value. Changing the onset value altered whether the fundamental increased in value at syllable onset or decreased; a low starting value resulted in a rising fundamental, and a high starting value resulted in a falling fundamental.) The authors asked subjects to rate the goodness and identity of the items, using a 9 point scale from -4 (an exaggerated example of /k/) to +4 (an exaggerated /g/). They used the data from this experiment to compute four perceptual measures: the subjects' category boundaries, the extent to which the subjects used pitch differences in making categorical distinctions, the extent to which they used the VOT, and the tradeoff between these two cues. The authors then asked each subject to produce the items /kIl/, /gIl/, /bIl/, and /pIl/ ("kill," "gill," "bill," and "pill"), and measured the subjects' mean VOTs for the voiced and voiceless items, the VOT differences between the two categories, and the differences in fundamental frequency at onset between the two categories. (Like /k/ and /g/, the primary difference between /p/ and /b/ is in voicing.) The authors found that while a number of perceptual measures correlated with one another, there were no correlations between any of the perceptual measures and their corresponding production measures.

There are three potential problems with this experiment that may explain this null result. First, differences in category boundaries between individuals tend to be quite small. In our laboratory, most voice onset time (VOT) series tend to only show individual differences in the range of one stimulus item or so (about 5-10 ms VOT). As the stimuli in this experiment only consisted of five different VOT values, it is quite likely that any differences between individuals would be too small to detect. One possible way to avoid this problem would be to use stimuli that had smaller inter-stimulus differences (that is, to make more items in the series). Another would be to use a measure that is more sensitive to slight variations in perception.

A second reason why this study may have failed to find a correlation lies in the VOT measurements the experimenters used. They averaged over the items "kill" and "pill", and over "gill" and "bill". Labial (/p/, /b/) and velar (/k/, /g/) stops tend to have rather different VOTs (Lisker & Abramson, 1964). These differences would likely mask the relatively small differences that might be expected between subjects. Any effect which might be present would be easier to find if production measures for different places of articulation were kept separate.

A third explanation for the lack of an effect in this experiment is based on the perceptual stimuli the authors used. Synthetic speech stimuli may not contain all the correlated cues listeners normally rely on when making categorization decisions, so this type of stimulus may not provide the best referent for actual speech perception. This is especially true as our ability to create synthetic speech has improved tremendously in the last decade or so. During the time period in which this study was performed, synthetic speech was not as high in quality as is currently available.

In a second experiment, Bailey and Haggard (1980) searched for a perception-production link in 2-year-olds. They synthesized five VOT series: bin-pin, bear-pear, deer-tear, goat-coat, and girl-curl. Children pointed to the appropriate picture for each word, and from this the authors found the children's perceptual boundaries (where responses were 50%

voiced) and the extent of ambiguity (the range between the points where the stimuli were labeled 20% voiced and 80% voiced). The children were also asked to name the items a number of times, and their VOTs were measured and averaged for each intended word. The authors looked for correlations between measures of perception and production, both for mean values and for measures of variability (standard deviation of produced tokens and slope of the perceptual function). Children whose voiced productions were at longer VOTs had perceptual boundaries that were paradoxically at shorter VOTs.[11] Similarly, the beginning of these children's region of ambiguity also began at a shorter duration. There was a trend for children who required longer VOTs perceptually in order to identify items as voiceless to also produce longer VOTs on these items, but this was not significant. This latter result is more in line with the idea of a link, but since it was only a trend no firm conclusions can be made. Furthermore, the significance of the negative correlation between voiced production and the category boundary makes it unclear how to interpret any of these results. There was a positive correlation between the slope of the identification function and the measure of productive consistency (standard deviations), suggesting that the degree of variability in both

---

[11] Note that this is the same result as that found by Raaymakers and Crul (1988) with misarticulating children.

measures are linked. However, this may be a maturity factor. That is, some children may be more mature, and thus more consistent in both tasks, whereas other children are more variable. There are a number of studies that have demonstrated more variability in children's productions than in adult productions, including two studies already mentioned above in the discussion on perception-production links in clinical populations (Lehman & Sharf, 1989; Smith, 1992). Unfortunately, then, this study seems to add as much confusion to the literature as it resolves.

A year after Bailey and Haggard's first null finding, Zlatin (1974) reported results described as supporting a perception-production link. She gave 20 adult subjects 4 synthetic speech series (bees-peas, bear-pear, dime-time, and goat-coat), each consisting of the central 15 members of what had originally been a 38-member series. These series, then, had far smaller differences between members than did the series used by Bailey and Haggard, which might explain the different results. Subjects were asked to identify the initial phoneme, and the author then calculated four different perceptual measures for each subject: the boundary location (the point at which the item was identified as voiced and voiceless equally often, or the 50% point), the upper and lower limits of the boundary region (the points at which the item was labeled with the voiced endpoint 75% of the time and 25% of the time), and the widths of the boundary region (the difference between the 75% and 25% points). The subjects were also asked to produce the eight test words, and these utterances were used to determine 6 different production measures for each subject: the average VOT (voice onset time) for voiced items, the mean lead time for voiced items (or average pre-voicing), the average VOT for voiceless items, the mean lag time for voiceless items, the range of productions (the difference between the highest and lowest VOT intervals used), and the discreteness of voicing categories. (It is unclear how the mean lag time for voiceless items is different from the VOT for these items. By most definitions, these two measures should be identical, and Zlatin does not describe her measures in enough detail to determine the difference.)

Zlatin then determined that 97.6% of the subjects' productions were within those subjects' perceptual phoneme categories. She uses this correspondence to argue that there must be a link between the perception and production. However, this may not necessarily be the case. She also found that while there was variation among subjects, the variation was not significant. Perhaps, then, humans just have a range over which their production can vary, and this range tends to be in the same range as their perceptual categories. This makes ecological sense: For communication to take place, a speaker's productions must be correctly interpreted and this requires that any given production fall within the correct category of the listener. Thus, there exists an ecological value to making sure one's productions are likely to fall within the intended category for any given listener. While there may be differences in production across individuals, these differences will be relatively small, so that all productions will still fall within the range that are likely to be correctly interpreted. And, to the extent that individuals try to produce tokens that will fall within the correct category of the listener, they will likely fall within the correct category of the producer, as well. Zlatin's results are necessary if there are limits on the extent to which productions can vary; that is, if people produce tokens intending for them to be correctly interpreted. Certainly, this does require some sort of connection between information in production and perception. Communication would never have developed without this sort of perception-production correlation, and most researchers would never argue against the existence of such a correlation. What is more debatable is whether there are consistent production differences between individuals within the range that would fall in the correct category, and whether these differences might be correlated with differences in perception in these individuals. This would suggest a much stronger link between the representations in production and perception than is suggested by the research described here, and might support the more general notion of a single processing mechanism that is involved in both production and perception. But this cannot be tested without finding a correlation between each individual subject's perception and production measures. Finding that when an individual wishes to produce a "p" he in fact produces something that sounds like a "p" is not sufficient to test this hypothesis.

Fox (1978; 1982) did test this hypothesis. He asked 16 subjects to perceptually scale the vowels /i, I, ᴇ, æ, ɑ, ʌ, o, U, u/ spoken by 6 speakers (five of the listeners were later dropped from the analysis, for inconsistent responses across trials). He then used INDSCAL to find the dimensions the subjects used in their scaling. He found 3 dimensions, which seemed to represent the height of the second formant (which corresponds to how far forward the tongue is during production, or how "front" the vowel is), the height of the first formant (corresponding to the height of the tongue during production, or how "high" the vowel is), and the presence or absence of lip rounding (which is commonly found in the English back vowels, such as /o/ and /u/, but not in the English front vowels). Although all of the remaining eleven subjects seemed to use the same three dimensions, they differed in the weightings (or saliences) they gave to each one. So Fox did a stepwise multiple linear regression to examine the relationship between each listener's perceptual weightings (his or her utilization of the different dimensions) and the acoustic measures of his or her productions. Fox used seven different sets of acoustic measures: F1 and F2 of the corner vowels (/i, u, ɑ/ - the vowels that are most extreme on the front-back and high-low dimensions) either with or without F0 (the speaker's fundamental frequency, or voice pitch), F1 and F2 of the non-corner vowels (/æ, ʌ, o/) including and excluding F0, F1 and F2 for all 9 vowels, plus F0; F1 values alone for all 9 vowels, and F2 values alone for all 9 vowels. He found that corner vowels are better predictors than non-corner vowels for the first two dimensions (F2 and F1 height), but not for rounding (the third dimension). More specifically, the F2 in production of /i/ and /u/ (the most extreme F2 values) were the best predictors for the F2 in perception, and the F1 of /ɑ/ and /i/ productions (the 2 extreme values of F1) were the best predictors for F1 perceptually. Fox argued that these correlations suggest that a perception/production link exists, and that it occurs at the level of phonetic classification.

This result is very suggestive. But the statistical analysis makes it unclear whether the results from these 11 subjects would generalize to the population at large. First, stepwise regression is designed to select from a group of independent variables the one which has the largest correlation with the dependent variable, and to test that particular correlation for significance first. This is contrary to hierarchical multiple regression, in which the investigator has an *a priori* reason to believe a certain correlation is the most likely, and thus tests the significance of that one prior to seeing if the others add any additional information. As Cohen and Cohen (1983) point out, the primary problem, then, with stepwise regression is that the significance test of an independent variable's contribution to predictability "proceeds in ignorance of the large number of other such tests being performed at the same time" and thus that such tests "can be very serious capitalizations on chance. The result is that neither the statistical significance tests for each variable nor the overall tests . . . at each step are valid." (p.124). This does not mean that such a correlation was not present within the 11 subjects tested, but it casts doubts on the likelihood of this finding being generalizable to the population as a whole.

Furthermore, Fox may not have been using the best acoustic measures. None of his acoustic measures are known to be especially relevant to lip-rounding, for example. In addition, the formant frequencies of vowels tend to change over the course of the segment. Without taking into account these changes, his measures may not be highly correlated with the acoustic cues people are actually using.

Lastly, all of the studies above (not just Fox's) have focused on the boundaries between phonological categories. But what is truly important for perception is really the category itself, not its boundaries. Boundaries are only indirect measures of categorization, at best. Since individuals' productions are unlikely to be ambiguous (or boundary) cases, these experimenters are forced to look for a correlation between the center of a production category (what the individual actually produces) and the boundary of a perceptual category. Using a perceptual measure that was also based on the center of the category would presumably be far more likely to show correlations with individuals' productions.

Paliwal, Lindsay, and Ainsworth (1983) attempted this. Like Fox, they used vowels in a /hVd/ context for similarity scaling. However, their stimuli were synthetic, rather than naturally spoken. This allowed the authors to vary F1 and F2 experimentally, creating a matrix of 12 F1 by 16 F2 frequencies (or 192 stimuli). Responses from each subject were used to determine the area in an F1-F2 space that corresponded to each of 11 different vowels, and to find the centroid of each area. The authors considered this centroid the prototype. The subjects then recorded the 11 possible /hVd/ syllables, and their F1 and F2 values were measured for each syllable. The authors compared the within- and between-subjects correlations on these measures. Presumably, a larger within-subjects correlation would suggest that there are links between the perception and production of each individual that are greater than what would be expected by chance (the between-subjects correlations). However, the authors found that the within-subject correlations were never significantly greater than the between-subject correlations (at a .01 level), although there was a nonsignificant difference for 9 of the 11 vowels in F1, and for 3 of the 11 vowels in F2. Transforming F1 and F2 from Hz to barks or mels (which are thought to be more representative of the scaling performed by the peripheral auditory system than are linear scales) did not alter this null result. The authors conclude that there is no evidence for a perception-production link.

Ainsworth and Paliwal (1984) extended this earlier study on vowels by examining F2 and F3 in the English glides (/w, r, l, j/). They varied the onset frequency of these formants in synthetic CV (consonant-vowel) stimuli, creating an F2/F3 space (10 values of F2 onset, and 10 values of F3 onset, for 100 total stimuli). They asked subjects to identify the initial consonants in these stimuli items, and also to produce tokens of these four syllables (/wE, rE, lE, jE/, or “weh”, “reh”, “leh” and “yeh”). The authors then measured these same formant-onset values for the subjects productions. As in the earlier study, they compared within-subject correlations and between-subject correlations, and found no significant differences (although there was a trend for higher within-subject correlations for /j/ and /r/ in F2 locus, and for /j/ and /l/ in F3 locus). Again, transforming the frequencies into barks or mels made no difference, and the authors concluded that there was no evidence for a perception-production link.

There are a couple of difficulties with this conclusion. First, there were only 10 subjects each for the vowel and glide experiments, and correlational results can be quite variable with small values of *n*. Also, it is uncertain whether they actually managed to find subjects' prototypes. As the 192 items in the first study were based on all pairings of 2 different dimensions, each dimension had a relatively small number of different values. Given that these different values were intended to be appropriate for the entire range of vowels, not for just one or two, they may not have had a sufficiently fine-grained series with which to find individuals' prototypes. The study with glides was somewhat better, but there were still only 100 items, consisting of 10 different values in each of the 2 dimensions, representing all 4 glides.

Lastly, both these studies, and the experiments by Fox, used simple measures of F1 and F2. However, recent work has suggested that these may not be the measures that are perceptually real to listeners. Two measures which have been suggested as being used by listeners are the differences between formants and the spectral moments of the signal (Syrdal & Gopal, 1986; Forrest, Weismer, Milenkovic & Dougall, 1988;

Sawusch & Dutton, 1992). Perhaps using either of these, more abstract measures, would result in a different pattern of results.

One final study examined the issue of productive representations in a different manner. Johnson, Flemming, and Wright (1993) asked listeners to select the best example of a given vowel from a set of 330 synthetic stimuli. These 330 items consisted of 15 different values of F1 and 22 different values of F2, and were intended to represent the entire vowel space. Thus, their space was slightly more fine-grained than that of Paliwal and Ainsworth (1983). Participants were also asked to produce tokens of the words “heed”, “hid”, “aid”, “head”, “had”, “H.U.D.”, “odd”, “awed”, “owed”, “hood” and “who’d”, and measurements of F1 and F2 were taken. However, rather than try and relate individual subjects’ perception and production, the authors looked at the averages across participants, and found that the vowel space was expanded in perception relative to production. That is, listeners expected (or preferred) to hear tokens that were outside the range of normal production. The authors suggest that underlying representations for productions reflect hyperarticulated versions of the vowels, rather than the vowel qualities found in more casual speech. Admittedly, even if all listeners prefer more extreme vowels than they actually produce, it does not necessarily mean that there could not also be production-perception links. That is, those individuals whose vowel spaces are most condensed (whose productions are least extreme) could prefer more moderate vowels, whereas individuals with relatively more extreme productions could prefer even more exceptional (even unnatural) versions. However, to the extent that listeners perceptual prototypes do not match anything found in normal production, the entire notion of perception-production links is called into question. It is hard to imagine how such results could fall out of system which used the same (or a similar) representation in the two modalities.

Thus, the situation remains unclear. The results do not seem particularly encouraging for the notion of a perception-production correlation. Certainly, such a link does not seem particularly robust. On the other hand, there have been a large number of studies that have found such a connection, hinting that there may really be some phenomenon worth investigating.

Most of the studies that have failed to find a correlation between perception and production have used synthetic speech with relatively coarse-grained distinctions between stimuli (for example, Bailey & Haggard, 1973; Bailey & Haggard, 1980; Ainsworth & Paliwal, 1984; Paliwal et al., 1983). In addition, some have averaged productions across different consonants (Bailey & Haggard, 1973; Bailey & Haggard, 1980), and others have used relatively simplistic production measures, such as individual formants (Ainsworth & Paliwal, 1984; Paliwal et al., 1983). Perhaps these acoustic measurements are not exact enough to provide consistent results. In support of this possibility, the only experiment that measured articulation directly (rather than measuring the acoustic properties that resulted from it) did find evidence of a perception-production link (Bell-Berti et al., 1979). This suggests that inconsistencies in other studies may be due, at least in part, to the measurement of acoustic properties.

There does not seem to be any consistent similarities between experiments which focused on the same set of phonemes. Although many of the successful studies have focused on vowels (Fox, 1978; 1982; Bell-Berti et al., 1979) others have used vowel stimuli with less success (Paliwal et al., 1983). Similarly, although many studies involving VOT have failed to find evidence of a link (Bailey & Haggard, 1973; Bailey & Haggard, 1980; Flege & Schmidt, 1995), others have had success using this cue ***(Flege, personal communication, 1996) (Hoffman et al., 1984; Flege & Eefting, 1986). Thus, it does not appear that the effects can be related to the sounds or acoustic features chosen as a basis of study. Nor can the variability be entirely explained by the specific task. Although all of the studies that have attempted to train listeners on perceptual cues have found concomitant improvement in production (Griffiths & Johnson, 1995;

Jamieson & Rvachew, 1992; Bradlow et al., 1997), similar consistency has not been found when the task involved correlating discrimination and production results (Travis & Rasmus, 1931; Kronvall & Diehl, 1954; Stitt & Huntington, 1969; Monnin & Huntington, 1974; and Hoffman et al., 1984; versus Mange, 1960; Prins, 1963; Haggard et al., 1971; Weiner & Falk, 1972; Raaymakers & Crul, 1988; Bailey & Haggard, 1973; 1980; Ainsworth & Paliwal, 1984; Paliwal et al., 1983). This suggests that some tasks may be more likely to find evidence of perception-production links than are others, but that task differences alone cannot account for the variability in the literature. Rather, results seem to depend on both the specific methodology used and the acoustic properties measured.

This variability suggest that there may actually be several factors which would need to be addressed in order to find perception-production links. To consistently find individual differences, a researcher would really need to make two correct decisions: He or she would need to choose a correct acoustic correlate as a production measure, and would need to choose a correct perceptual task. These decisions are not as simple as they might seem, and in fact, most of the studies above did not succeed at them. Choosing a correct acoustic correlate is difficult for several reasons. To begin with, we still do not actually know what perceptual dimensions speakers use when listening to speech. Furthermore, one of the classic difficulties in speech perception research is the problem of invariance. There does not seem to be any single cue which occurs in all instances of a given phoneme. That is, the same sound will be produced differently in different contexts and by different speakers, and thus there does not appear to be any cue which is invariant across the different tokens of an intended phoneme. So, choosing an acoustic measure is not simple. But unless the researchers choose a measure that is at least highly correlated with the perceptual dimension the listeners are using, it would be very difficult to find any perception-production links on the basis of that measure. That is, if the researchers choose to study an acoustic measure that is not strongly related to the cue the listeners are actually using, it is highly unlikely that the researcher would be able to find any suitable differences between subjects.

In addition to choosing a cue that is correlated with what the listener uses, the researcher needs to choose a proper task. Several of the studies discussed above have used perceptual tasks that focus on the boundaries between phonological categories. But what is truly important for perception is the category itself, not its boundaries. And, as stated earlier, since individuals' productions are unlikely to be ambiguous (or boundary) cases, these experimenters are forced to look for a correlation between the center of a production category (what the individual actually produces) and the boundary of a perceptual category. Using a perceptual measure that was also based on the center of the category would presumably be far more likely to show correlations with individuals' productions.

There has recently been some work which provides a new way of examining the centers of perceptual categories. Miller and Volaitis (1989) created a VOT series which ranged from a clear /bɑ/, through some stimuli that were ambiguous between /b/ and /p/, to a clear /pɑ/, and then beyond the good /pɑ/. (The authors labeled these extreme stimuli as /*pɑ/.) These extreme stimuli sound like a very breathy "pa", and have voice onset times that are far larger than would normally occur in speech. Subjects were asked to rate each of the items as to how good of a "p" they were. As expected, the very short stimuli were heard as /b/, and thus received very low ratings. As the stimuli became more /p/-like, their ratings increased. But, as the voice onset time became too long for a typical /p/, the ratings dropped again. Not only did subjects' ratings drop for the extreme items, they also showed different ratings even among the good exemplars. Usually, only one or a few items received the very highest ratings, even though the neighboring items might still be heard as good examples of the category. That is, even among those items that the subjects would have labeled as being a "p" (rather than some other phoneme), or even a decent "p," the tokens still varied in their goodness. This suggests that experimenters can measure not only the boundaries between categories for each subject, but also can measure the subject's organization within any given category.

Thus, Miller and Volaitis (1989) have shown that prototypes (or best examples) for phonetic categories do exist, and can be measured by using an appropriate task. This type of task might be expected to provide a better perceptual measurement for perception-production correlations than would a task based on boundary measures. While there were a few studies discussed above that attempted to find these prototypes (Paliwal et al., 1983; Ainsworth & Paliwal, 1984), the differences between their procedures and those of Miller and Volaitis may have made this impossible. Miller and Volaitis used approximately 40 items, varying in only one dimension, in order to find the prototype for just one phoneme. This is in contrast to the studies by Paliwal and Ainsworth, which used only 10-16 items per dimension, and used these to find prototypes for several different phonemes. In fact, of the 36 items in Miller and Volaitis' series, only 6 were rated above an 8 (on a 10-point scale). Thus, ratings began to drop off quite quickly as the stimuli moved away from the prototype. This might suggest that unless a sufficiently fine-grained series was made, the prototype might be missed altogether. The Ainsworth and Paliwal study was slightly better in this regard, having 100 stimuli consisting of only 4 phonemes, but these 100 stimuli differed in two different dimensions. It is possible that 10 values of each formant may still may be too few to find an accurate measure of the perceptual prototype. Since differences between subjects are likely to be small, only a task that is sensitive to slight differences in prototype locations across subjects could reasonably be expected to produce a measure that would correlate with differences in production. This might help explain the null results in the studies by Ainsworth and Paliwal.

With these problems in mind, I decided to make a more sensitive test of the existence of a production-perception link. The perceptual task was modeled after the one used by Miller and Volaitis. This task examines the centers of phonemic categories, not the boundaries between them. Also, the very small differences between stimuli in this study should make it possible to find differences between subjects as to the location of phonemic prototypes. In order to avoid choosing a poor acoustic correlate, this first experiment is based on VOT differences between voiced and voiceless stop consonants. This acoustic measure is well-known to be correlated with the dimensions humans actually use in perception (Lisker & Abramson, 1964; Lisker & Abramson, 1970). Furthermore, VOT values do not appear to differ significantly with talker dialect (Syrdal, 1996). This is important because differences in dialect would already be expected to appear in both perception and production, regardless of the presence of a direct link. As we wish to examine perception-production correlations within a dialect, using a cue known to be dialect independent prevents possible confounds. Finally, there has been some suggestion in the literature that VOT values in production and perception might correlate (Hoffman et al., 1984).

CHAPTER 2

Experiment 1: The production and perception of VOT

This experiment was designed to investigate whether individuals' perception of speech contrasts is linked with their production of those contrasts. In order to determine this, listeners were asked to participate in both a production and a perception task, and then correlations were calculated between each subject's measures on the two tasks.

For the production task, a female speaker (RSN) recorded three tokens of the target syllable "pa" (/pɑ/), and one token each of the other CV (consonant-vowel) syllables consisting of the 6 English stop consonants (/p/, /b/, /t/, /d/, /k/, and /g/) and the vowels /i/, /e/, /æ/, /u/, /o/, /ɔ/, /ʌ,/, and /ɑ/ (the vowels that occur in "beet", "bait", "bat", "boot", "boat", "bought", "but" and the second syllable of "robot"). These vowels were chosen because they represent the entire range of monothongal vowels that occur in English, and because all of them could occur in an open syllable (that is, in a consonant-vowel, or CV environment). The subjects heard these syllables one at a time over a loudspeaker, and were asked to repeat back each syllable in the way that they would normally produce it (that is, they were not supposed to mimic the speaker, but to produce the utterances naturally). These recordings were stored for later acoustic measurement.

The perception task was modeled on work by Miller and Volaitis (1989). They created a VOT series which ranged from a clear /ba/ to a clear /pɑ/, and beyond a good /pɑ/ (/*pɑ/). They presented these stimuli in random order to their subjects, and asked the listeners to rate how good an example of the category /p/ each stimulus was. Miller and Volaitis considered the highest ranked stimulus to be the listener's prototype for the category /p/.

This prototype measure is likely to be very sensitive to individual differences between subjects. For this reason, the current experiments are modeled on the procedure Miller and Volaitis used. Specifically, a series was created that ranged from /bɑ/ to /pɑ/ to /*pɑ/. Subjects heard these stimuli in random order, and were asked to rate the stimuli as an example of the item "pa".

If the VOT for a listener's prototypical "pa" in the perception task is correlated with the VOT that an individual listener produces in the production task, it would suggest that there is a link between the perceptual and productive aspects of speech. Failure to find this correlation would throw into question the various models which require such a link.

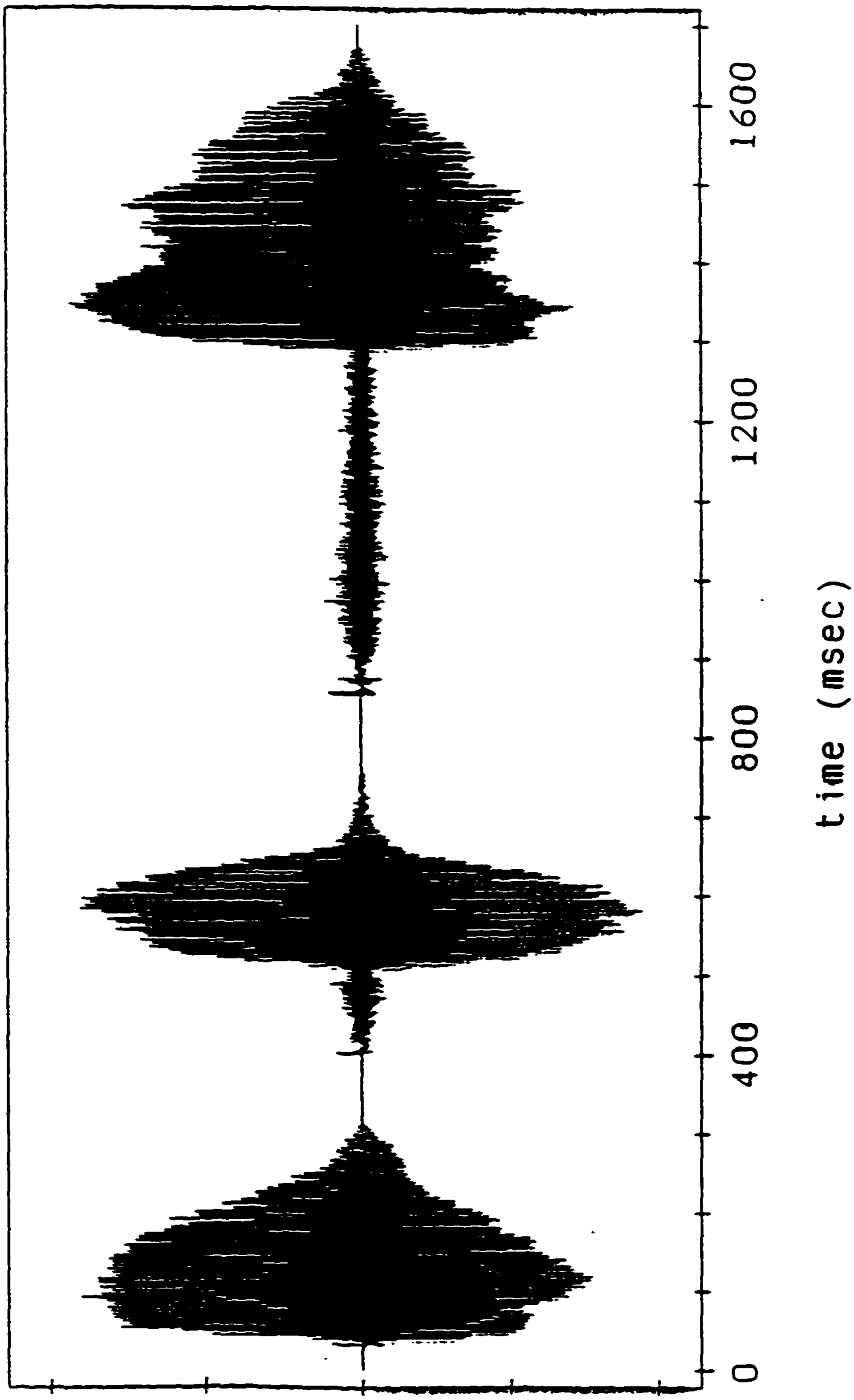
Method

Subjects. The listeners were 27 volunteers from the Buffalo community. The participants took part in 2 one-hour sessions within the same week, and received $10 in compensation at the end of the second day of the experiment. All were native speakers of English with no reported history of a speech or hearing disorder. During debriefing it was discovered that one of our listeners was not a native speaker of English; his data were not included in the analysis. An additional subject missed her second appointment. Her data were likewise not included. A further five subjects were dropped from the experiment because a central member of the /pɑ/ category could not be determined from their perceptual data. These subjects' ratings did not drop for even the most extreme values of VOT. That is, they rated items with VOT values over 200 ms long as highly as they rated items within the range of 50 to 150 ms (where the other subjects' prototypes lay). It is possible that these subjects did not understand the instructions, and were simply rating the stimuli as to whether they were a "p" or not, rather than rating them as to their category goodness. Or, perhaps these subjects did understand, and were merely outliers. These subjects may, in fact, have been demonstrating a hyperarticulation effect in perception of stop consonants analogous to that found in vowels by Johnson *et al.* (1993). Regardless, including their data would have masked any effects of individual differences that were present. Leaving out these listeners resulted in 20 participants for this experiment.

<u>Stimuli.</u> For the production task, a female native talker of English (RSN) recorded one token each of the 48 CV syllables formed from all possible pairings of the six English stop consonants (/p/, /b/, /t/, /d/, /k/, /g/) and the eight vowels /i, e, æ, u, o, ɔ, ɑ, ʌ/. Two additional tokens of the syllable /pɑ/ (for a total of three) were recorded to provide a greater range of examples of the target syllable. All of the tokens were amplified, low-pass filtered at 9.5 kHz, digitized via a 16-bit, analog-to-digital converter at a 20 kHz sampling rate and stored on computer disk.

For the perception task, the same native speaker recorded the tokens /bɑ/, /pɑ/ and /*pɑ/. Figure 1 shows waveforms of these three items. Here, time is on the x-axis, and amplitude is on the y-axis. During production of the syllable /pɑ/ (or /bɑ/), a speaker closes his lips and allows air pressure to build up inside the oral cavity. Once pressure is sufficient, he opens his mouth. A burst of air rushes out, creating a "noisy" sound. (This puff of air can be felt by placing your hand in front of your face and saying "puff"). In Figure 1, this sudden release burst is the sharp peak at the very beginning of the syllable. This is followed by a low-amplitude, irregular section representing the noise.

At some point following the release burst, the speaker's vocal folds begin vibrating. This creates the more regular pattern of vertical lines in


0
400
800
1200
1600
time (msec)

the higher-amplitude portions of Figure 1. Each "line" represents one cycle of the vocal folds opening and closing.

The distinction between a "b" and a "p" is in the time-delay between the release of air pressure in the burst, and the onset of vocal fold vibration. For a "b", this time delay is typically quite short, on the order of 0 - 5 ms. Oftentimes the vocal fold vibration even begins prior to the burst. (This is known as "prevoicing", and will be discussed again later.) For a "p", there is typically a delay of 40 to 120 ms before the onset of vocal fold vibration. Although there are some acoustic differences between the noise portion of a /b/ burst and that of a /p/ burst, the duration of this aspiration is considered the primary cue differentiating these sounds. For the /*p/ tokens in this experiment, the delay was extended beyond this normal range. This is apparent in Figure 1. In the /b/ production (to the left) there is almost no delay between the first burst and the beginning of vocal fold vibration. In /p/ there is a much longer noise portion (about 1 cm long in the figure, representing approximately 107 ms). In /*pɑ/, the noise portion is even longer, well over an inch in the figure, or over 400 ms.

A 21 item continuum ranging from /b/ to /p/ was created from the /ba/ base by removing successively longer sections from the /b/ onset and replacing them with the corresponding sections of the /p/ (/pɑ/) onset. This serves to replace more of the vocal fold vibration of the /b/ with the aspiration (or noise portion) of the /p/, and creates a series with successively longer portions of aspiration. The editing procedure used to produce these stimuli is essentially identical to that used by Miller and Volaitis (1989). The first stimulus was created by removing the /b/ release burst at the onset of /ba/ (8.4 ms) and replacing it with the release burst from /pa/, resulting in a stimulus with the same VOT as the original /b/ (that is, the same noise duration) but with the release burst of a /p/. All editing was done at zero crossings in the digital waveform to avoid audible clicks or other distortion. The second stimulus was made by removing the /b/ burst and the first vocal pulse, and replacing this with the equivalent /p/ burst plus aspiration duration. The third through twenty-first stimuli were each made by removing one additional vocal pulse from the onset of the /ba/ syllable than did the prior stimulus. These vocal pulses were replaced with the equivalent duration of burst release and aspiration. The durations of the vocal pulses were not exactly equal, but averaged 4.17 ms. Then, 40 additional items were generated. Here, aspiration was removed from the /*pa/ token and added to the end of the aspiration in the last item of the /b-p/ series (i.e., the twenty-first, or most "p"-like item). Each successive item contained approximately 5 ms more aspiration than did the item before. In these stimuli, the number of vocal pulses remained the same as the number in stimulus 21. That is, as additional aspiration duration was added beyond a VOT of 90.5 ms, the duration of the voiced portion of the syllables was held constant.

This resulted in a 61 item series, which would have taken a bit too long to run. However, it was necessary to maintain the small VOT differences, in order to be sensitive to small differences in prototypes between subjects. Pilot testing was used to determine the range of stimuli over which most listeners' prototypes fell. It was found that most individuals placed their prototypes between 55 and 140 ms VOT (or between stimulus items 13 and 31). Therefore all of the stimuli within this range were included in the experiment. Beyond this range, every other stimulus was included in the experiment, and the remaining stimuli were removed. This resulted in a 40-item series, with VOT differences of 4.6 ms at intermediate VOTs and 9.4 ms at both longer and shorter VOTs. The VOT values for each of these 40 items are shown in Table 1.

Procedure. Listeners were run individually, in two separate sessions, and participated in the production study at the beginning of the first session.[12] For the production study, the subjects were seated in front of a Digital Equipment Corporation VAX station 4000 computer, which

---

[12] Because of a computer error, one subject's production task had to be recorded at the start of the second day's session.

Table 1

VOT values for members of the /ba/ - /pa/ - /*pa/ series used in Experiment 1.

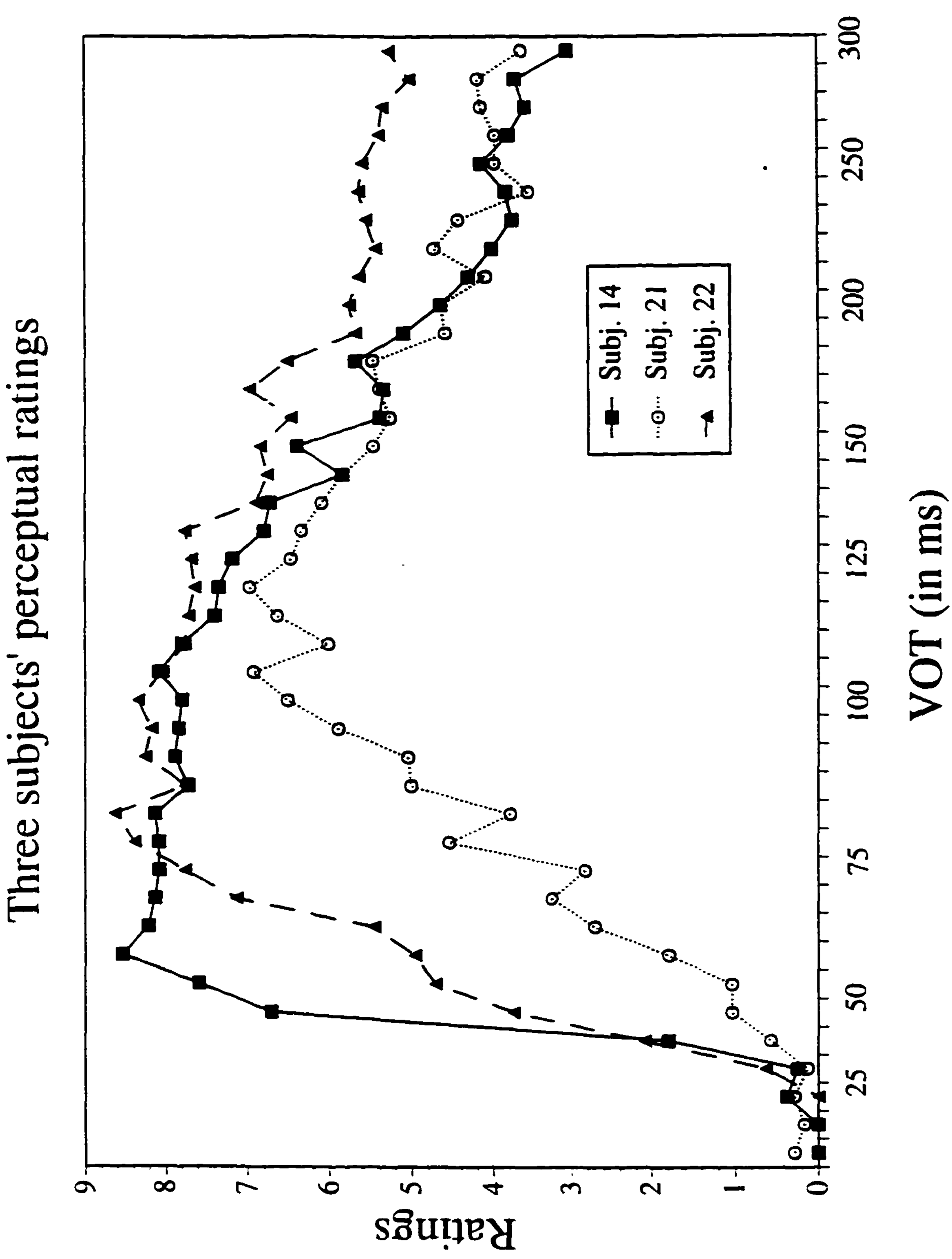
Item #	VOT	Item #	VOT
1	8.25	21	120.90
2	15.00	22	125.60
3	23.15	23	130.70
4	31.65	24	135.60
5	40.15	25	140.80
6	48.60	26	150.85
7	57.25	27	160.60
8	60.85	28	171.20
9	65.60	29	181.70
10	70.00	30	190.70
11	74.40	31	200.60
12	78.15	32	210.60
13	81.85	33	220.55
14	87.95	34	230.35
15	90.50	35	240.65
16	96.00	36	250.55
17	100.05	37	260.55
18	105.30	38	270.95
19	110.15	39	280.90
20	115.65	40	291.00

controlled stimulus presentation and response collection. The subjects held an Electro-Voice D054 Dynamic Omni Microphone, and listened to the stimuli over a Realistic loudspeaker. The stimuli, which were stored on disk, were converted to analog form by a 16-bit, digital-to-analog converter at a 20 kHz sampling rate, and low-pass filtered at 9.5 kHz. The syllables were presented in random order. Listeners were asked to repeat each syllable into the microphone in the manner they would normally produce that syllable. The computer waited 4 seconds for a response. If the subject did not respond within that time frame, the computer presented an error message and presented that trial again. Also, if the subject's response was too loud (peak-clipped), the computer would similarly repeat the trial. Otherwise, the computer gave the listener the opportunity to decide whether or not to keep that trial. Subjects were instructed to respond "no" if they were unsure of what they were supposed to have said, or if some other noise interfered with the recording (for instance, a cough). If the subject responded "no", the trial was repeated. Otherwise, the program proceeded to the next trial. There were a total of 50 trials in this block. The program was then run a second time, so that each subject recorded two tokens of each CV syllable (and 6 tokens of the target item /pɑ/).

The subjects were then moved to a second experiment room, and seated in front of a Macintosh Centris 650 computer, which controlled stimulus presentation and response collection for the perception task. Again, the stimuli were converted to analog form by a 16-bit, digital-to-analog converter at a 20 kHz sampling rate, and low-pass filtered at 9.5 kHz. They were amplified and presented binaurally through TDH-39 headphones at a comfortable listening level. The syllables were presented in random order. Listeners were asked to rate the initial phoneme for its goodness as an example of the category /p/. Subjects responded using the numbers zero through nine on a numeric keypad, followed by the "return" or "enter" key. Subjects were told to use the "0" label whenever the item did not sound like a "p" at all, to use the "1" whenever it was unclear whether it was a "p" or not, and to use the range "2" through "9" for items which were definitely members of the category "p", but differed in how good of an example they were. They were given a reference sheet which contained this scale, in case they wished to refer back to it. While subjects' response times were not recorded, they were informed that the next trial would begin as soon as they responded to the current trial. Responses from the first block of trials (one repetition of each item) were considered practice, and were not included in subsequent data analysis. After the practice set, stimuli were presented in blocks of 80 trials (two repetitions of each stimulus). Listeners participated in six blocks of experimental trials in each of the two sessions, resulting in a total of 24 responses to each stimulus.

Results and Discussion

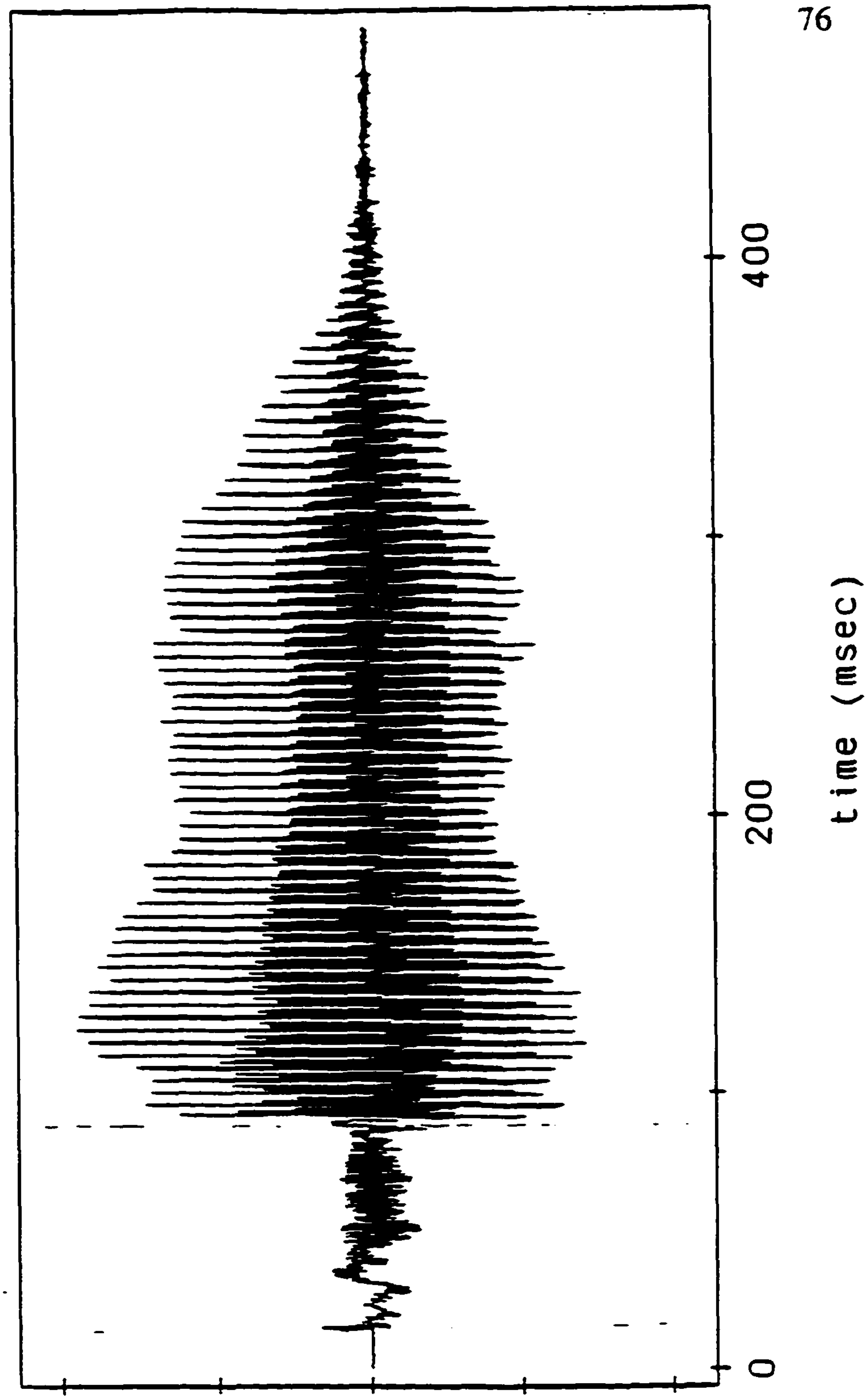
A mean rating was computed for each stimulus for each subject in the perception experiment. The single item with the highest rating was considered to be the listener's prototype. The VOT of this item was thus recorded as that subject's prototypical VOT for the item "pa". For five subjects, this item had a VOT of over 200 ms. The data of these subjects were removed, as described in the methods section above. One subject had equally high ratings for 2 items in the continuum. The VOT values for these two items were averaged to find that subject's prototype. Figure 2 shows the rating functions for three subjects who participated in this experiment. As is clear from this figure, the subjects' ratings tended to increase until they reached a peak, and then immediately began decreasing, leaving a single item as a prototype. For some subjects, two or three items with similar VOTs received quite high ratings, although usually one was slightly higher than the others. This highest item was treated as the prototype, even when it received only slightly higher ratings than another member of the series. It is unclear whether these slight rating differences represent actual perceptual differences or are caused by random


Three subjects' perceptual ratings
Ratings
0
1
2
3
4
5
6
7
8
9
VOT (in ms)
25
50
75
100
125
150
200
250
300
Subj. 14
Subj. 21
Subj. 22

fluctuations in ratings. If the latter, then treating the single highest item as the prototype may add additional noise to the data. However, this is likely to be relatively minor, as it was generally only nearby members of the series (with comparable VOTs) that had similarly high ratings. Furthermore, the subjects' prototypes varied over a relatively large range (60.9 to 150.9 ms). Even if the choice of a prototype was off by one, or even two, members of the series for each listener, the variability across listeners would still be present. Plus, there is a precedent in the literature for selecting the single, most highly rated item as the prototype (Miller & Volaitis, 1989).

For the production experiment, the time interval from syllable release to the onset of voicing was measured for each token produced by each subject. Figure 3 demonstrates how this measurement would be made. This is an example of one individual's production of "pa". Two dotted lines have been added to the figure. The first line is located at the onset of the burst, and the second is at the onset of vocal pulsing. The VOT is the duration between these two lines. In the example in Figure 3, the VOT is 73.55 ms.

The six values for the recordings "pa" were averaged to determine the produced "pa" VOT. The values for the 14 recordings of the other 7 "p" syllables (/pi/, /pe/, /pæ/, /pʌ/, /pu/, /po/ and /pɔ/) were averaged to


0
200
400
time (msec)

find a mean VOT for the remaining "p" productions. Likewise, the values for the 16 recordings for each of the other stop consonants were averaged, to determine a mean VOT for each stop consonant. For voiced items which were prevoiced (where the vocal folds began vibrating prior to release), the prevoicing itself was not included, because it may not be valid to average across values of prevoicing and aspiration. That is, if a subject prevoiced one instance of /be/ ("bay"), the duration from the release to the onset of the vowel was measured, and the prevoicing ignored. Had this not been done, the negative value of prevoicing would have been averaged along with the positive values of the tokens which were not prevoiced. As bursts and prevoicing are different cues, it may not be appropriate to combine them in this manner. In order to prevent any systematic bias on the part of the experimenter, the tokens were measured in the same random order as they were produced by the participant. Furthermore, the experimenter did not know the results from the perception experiment until the measurements were completed for the production experiment.

The methodology just described resulted in one perceptual measure (VOT of the prototype), and seven production measures (average VOT for /pɑ/, average VOT for the other /p/ items, and the average VOT for the /b/, /t/, /d/, /k/, and /g/ items). These values are shown in Table 2. To perform each of these correlations separately would have resulted in 7

Table 2

VOTs in production and perception (ignoring prevoicing)

	Production							Perception
Subject	pa	p	t	k	b	d	g	pa
ALG	78.0	70.0	75.0	81.2	20.8	24.3	35.6	70.0
BTK	74.4	67.0	75.6	77.1	15.3	21.2	31.0	78.2
CMA	53.8	55.1	65.0	65.0	8.7	15.2	21.1	81.9
DG	61.5	83.3	80.3	90.7	17.4	20.0	22.7	78.2
ETP	110.6	123.2	123.1	121.4	11.9	18.0	23.2	120.9
FNP	87.8	77.1	87.3	94.4	6.0	9.1	17.9	150.9
JMD	51.1	54.6	65.0	66.3	13.8	14.1	19.8	60.8
KAF	55.7	58.8	68.0	65.2	10.9	16.2	21.6	88.0
KLG	76.4	75.7	76.2	81.5	9.0	14.1	23.3	100.1
LCG	65.1	58.9	75.8	73.7	12.1	15.5	21.1	60.9
LEP	63.7	85.5	102.1	93.0	15.7	25.4	26.4	74.4
LMR	52.1	53.5	56.7	57.0	14.0	14.7	25.0	74.4
LMZ	93.7	88.2	91.2	88.4	12.4	19.1	26.2	88.0
PEG	74.9	72.6	98.1	89.7	9.5	23.4	20.9	89.1
SJC	61.9	66.9	92.0	92.1	7.1	19.9	13.8	96.0
SJD	70.5	89.9	102.9	101.0	8.5	17.2	23.5	65.6
SLD	73.8	72.9	101.3	100.7	16.4	20.4	28.4	105.3
TAH	124.8	114.8	117.4	116.4	10.7	22.7	25.9	105.3
TEB	66.6	71.1	85.0	83.6	17.5	30.2	27.4	74.4
TVK	91.4	81.9	87.4	98.3	10.7	20.6	33.2	90.5

different correlations. With this many analyses, the possibility of a spuriously significant finding is rather high. Rather than this, a hierarchical regression was performed, using the perceptual measure as the dependent variable, and all seven production measures as independent variables. A hierarchical regression requires an *a priori* ordering of the independent variables in terms of their likelihood of having a correlation with the dependent variable. Presumably the average VOT of /pɑ/ would be the most likely to show an effect, since it is the best referent for the perception of /pɑ/. The average VOT of the remaining /p/ items would be the next best referent, as they contained the same initial phoneme. It is less clear which item should come next. In general, the VOT of alveolars tend to be more similar to bilabials than are velars (Lisker & Abramson, 1964; Klatt, 1975; Dorman, Studdert-Kennedy & Raphael, 1977). Thus, the VOT of /t/ items should be higher in the listing than those of /k/, and the VOT of /d/ should be more likely to have a correlation than that of /g/. Furthermore, items which differ from the relevant consonant in only one feature should be more likely to show a correlation than do items which differ in two features. Thus, /t/ and /k/ should be higher in the hierarchy than should /d/ and /g/. But it is unclear whether /t/ or /b/ should be a better referent: /t/ shares the voicelessness of /p/, but /b/ shares the place of articulation. There are good arguments to be made for either ordering, as both items differ in only one feature. Because of this difficulty, the regression was performed twice, once with the ordering /pɑ/, /p/, /t/, /k/, /b/, /d/, /g/, and the other time with the ordering /pɑ/, /p/, /b/, /t/, /k/, /d/, /g/. While this increase does make a Type II error slightly more likely, it is believed that this risk is small enough as to be outweighed by the potential benefits. The results from these regressions are shown in Table 3.

It is important to note that a multiple regression searches for additional predictability. Because of this, an independent variable (IV) may be highly correlated with the dependent variable (DV), and yet contribute nothing to the regression formula. As an example, if a large effect were found for the VOT of /pɑ/, and no effect was found for the VOT of the remaining /p/ items, this would not necessarily mean that the VOT of the /p/ items was not correlated with the DV. Rather, it means that the inclusion of the second IV (/p/ VOT) did not add any additional information (or predictability) to the equation. This could occur anytime the IVs are themselves correlated. If the VOT for the /pɑ/ productions were correlated highly both with the DV (the perception measures) and with the VOT of the remaining /p/ items, the latter would necessarily be correlated with the DV as well. However, this correlation would not contribute to the regression formula. The regression only reports

Table 3

Results from multiple regression from Experiment 1

Ordering: pa, p, t, k, b, d, g

Step	Multiple *r*	Multiple r^2	Change in r^2
pa	0.5743 *	0.3299	0.3299 *
p	0.5802	0.3366	0.0068
t	0.5930	0.3516	0.0150
k	0.6420	0.4121	0.0605
b	0.7408 *	0.5488	0.1367 *
d	0.7778	0.6050	0.0562
g	0.7952	0.6323	0.0273

Ordering: pa, p, b, t, k, d, g

Step	Multiple *r*	Multiple r^2	Change in r^2
pa	0.5743 *	0.3299	0.3299 *
p	0.5802	0.3366	0.0068
b	0.6954 *	0.4835	0.1469 *
t	0.6973	0.4862	0.0027
k	0.7408	0.5488	0.0626
d	0.7778	0.6050	0.0562
g	0.7952	0.6323	0.0273

correlations between the IV and the DV once the correlations from the prior DVs have been partialed out.

The results from the regression with the ordering /pɑ/, /p/, /t/, /k/, /b/, /d/, /g/ will be considered first. All of the IVs were highly correlated with the DV, but only the VOT values from /pɑ/ and the /b/ items contributed significantly to the formula. The variation in produced /pɑ/ VOT was responsible for 33% of the variance in subject's perceptual ratings (*F*=8.86), whereas the variation in /b/ VOT was responsible for an additional 13.7% (*F*=4.24). A complete listing of the correlations, r^2, and change in r^2 is given in Table 3.

The results with the alternative ordering were similar. All of the IVs were correlated, but only the /pɑ/ and /b/ values contributed to the regression formula. The variation in /pɑ/ productions had the same value (as its place in the hierarchy was unchanged). The variation in /b/ productions was responsible for an additional 14.7% of the variation, after the prior items' correlations were considered.

These results suggest that there is a link between perception and production. There were significant correlations between each subject's productions and their perceptual prototypes. Thus, subjects whose prototype for "p" occurred at longer VOTs also tended to produce longer VOTs.

It is also interesting to note that while the listeners' productions of the voiceless stops did not provide any additional information above and beyond their productions of the target item itself, their productions of the first voiced stop in the hierarchy did. This suggests the possibility that production of voiceless items may be highly correlated within each individual, but that production of voiced items may not be as correlated with the voiceless tokens. That is, the additional voiceless stops may not have added additional information because they were highly correlated with the production of the target item. Since the first voiced stop did add additional predictability, it must have contained additional information about the talker beyond that provided by the production of the target item. To the extent that it provides different information, it is not highly correlated with the production of the voiceless items. To examine this further, a correlation matrix of the seven perceptual measures was performed. These correlations are shown in Table 4, and clearly show that the voiceless items were highly correlated with each other, and the voiced items correlated highly with one another, but the correlations between voiced and voiceless items were far lower.

The results from this experiment clearly show that individual differences in production are related to differences in perception. These

Table 4 Correlations among production measures

Correlations without prevoicing

	pa	p	t	k	b	d
p	.854					
t	.738	.881				
k	.801	.910	.954			
b	-.151	-.076	-.149	-.133		
d	.145	.232	.350	.278	.578	
g	.275	.151	.026	.098	.664	.515

production-perception correlations can be found if the researcher chooses an appropriate perceptual task and acoustic correlate.

A relevant question is why the correlation is not even higher. There are a number of sources of noise in the data that might have contributed to this. Individuals do not always produce tokens at the same VOT. That is, there is intrasubject variability, as well as intersubject variability. An average VOT measure across six productions was used as a way of accounting for this variability. But it is quite possible that 6 tokens was an insufficiently large sample size in this regard. Results from anchoring studies suggest that perception likewise varies over time, depending upon perceptual context, and this may have been a factor in the present experiment as well. Subjects heard each item in the perceptual series 24 times, which I hoped would have been sufficient to find a stable estimate of subjects' prototypes. However, this may have been too few trials. Furthermore, the perception task took place over two days, whereas the production task took place during only one. If perception does vary with time, perception during the second session may have been different than that during the first session. This would result in greater variability for the perceptual data.

Although step sizes were fairly small, it is still possible that none of the items in the series truly represented the subjects' prototypes. This, too, would have added noise to the data. Finally, with 100 productions for each of 20 subjects, some degree of measurement error and error in data entry is likely to have occurred.

However, all of these possibilities can only explain relatively small differences between perception and production (those caused by noise, rather than those caused by consistent differences). Examining Table 2, it is apparent that there are some sizable differences between the perception and production measurements for some participants. It is unclear why subjects such as KLG and FNP had such vastly different measures for their perceptual prototypes as for their own productions, but these differences are unlikely to have been due to noise alone. One possibility is that the perceptual responses for these listeners were affected by the range of productions present. Another is that they recognized the talker in the perception task as the one they had just heard during the production task, and judged the perceptual items relative to what they already knew about that talker. There is no obvious way of distinguishing between these different possibilities at this point, but it suggests that it might be prudent in future experiments to use different talkers when creating the stimuli for the perception and production tasks.

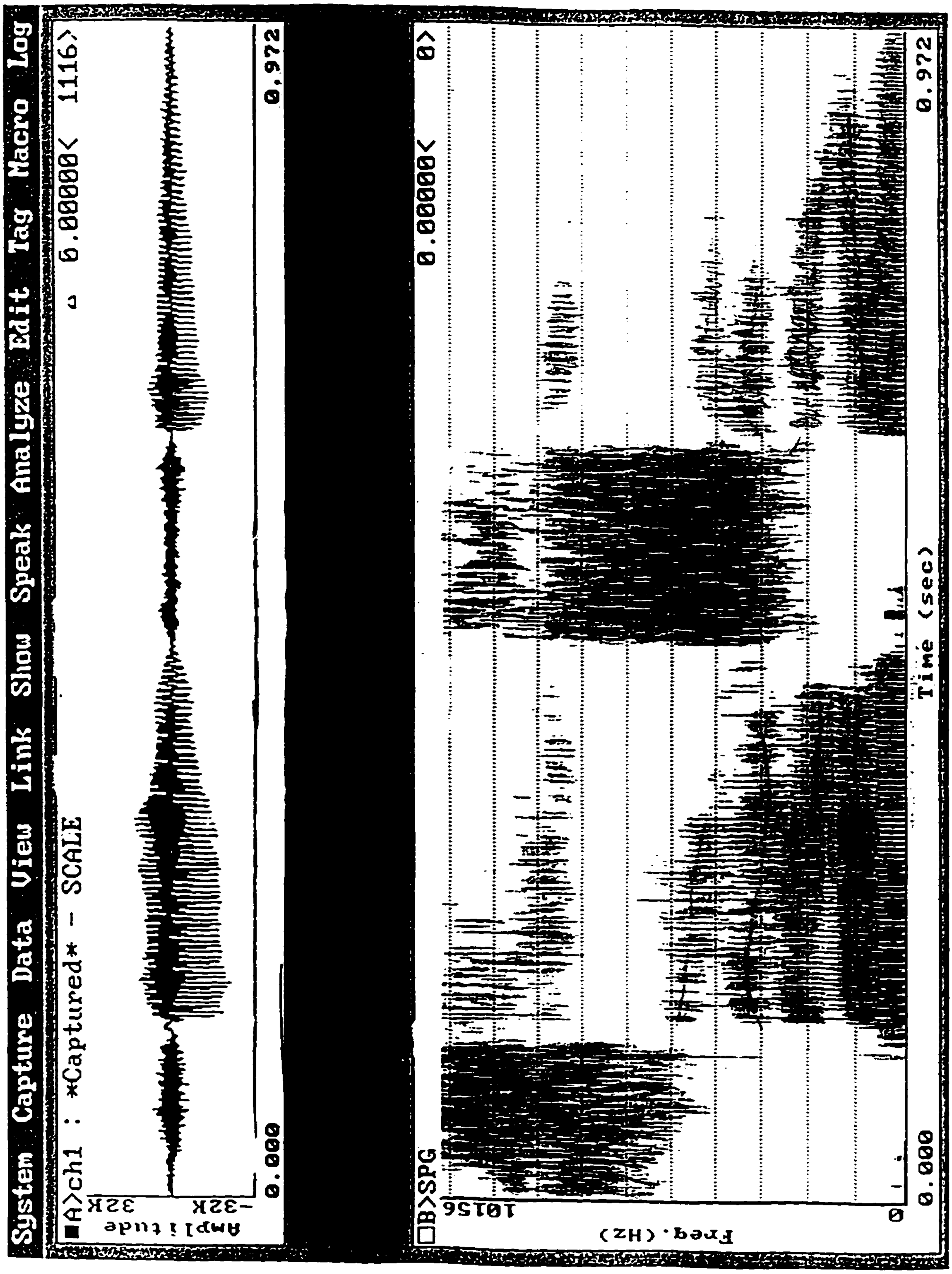
There is one further possible explanation for some of these perception-production differences. For the majority of the talkers with large differences between the VOT measures in the two tasks, the VOT in the perception task was larger than in the production task. This suggests that subjects may have preferred listening to a more extreme token of the sound than they actually produced themselves. Recent research has found these hyperarticulation effects for vowels (Frieda, 1997; Johnson et al., 1993), and it is not implausible that a similar process would occur for consonants, as well. In support of this possibility, the VOTs of voiceless /pɑ/ tokens across all subjects were reliably larger in the perceptual prototypes than they were in production ($t=3.11$, $p <..006$). Fourteen of the 20 participants demonstrated this pattern of larger VOTs in perception than production (that is, of preferring more extreme tokens of /pa/ than they actually produced). It is unclear why the other six participants did not show such an effect, or why the size of this effect varied across individuals. Perhaps some individuals demonstrate these hyperarticulation effects more strongly than do others.

Despite these differences, the present methodology seems to provide fairly consistent results across a group of subjects. One possible application of this task is that it could be used to evaluate different acoustic correlates. That is, we now know that an acoustic measure that is appropriate (i.e., one which is highly correlated with what the listener actually uses as his or her perceptual dimension) demonstrates correlations between production and perception. Given that, perhaps the appropriateness of a new acoustic measure can be evaluated according to whether or not it, also, shows such a correlation. Furthermore, one might suspect that a cue that was more closely linked with the perceptual dimension used by a subject would show a higher correlation between production and perception than one that was less intimately linked. The second experiment will test this possibility.

CHAPTER 3

Experiment 2: The production and perception of fricatives

This second experiment examines two different potential cues to the same phonemic distinction. An /s/-/ʃ/ ("s" - "sh") distinction was selected, with /ʃ/ as the primary phoneme of interest, because there appear to be two easily identifiable potential cues for this distinction. Both of these phonemes are voiceless fricatives. This class of sounds is produced by creating a partial obstruction in the mouth. Forcing air through this narrow constriction causes turbulence in the air-stream, resulting in a "noisy" sound. This noise consists of energy at a broad range of frequencies (Pickett, 1980). The obstructions are formed by the tongue pressing against the top of the mouth. However, this obstruction takes place further forward in the mouth for the /s/ than it does for the /ʃ/. The oral cavity in front of the constriction filters the noise, emphasizing and de-emphasizing certain frequencies. Different-sized cavities have different patterns of emphasis, with smaller cavities resulting in higher frequencies. Since the constriction is farther forward in the vocal tract for the /s/, there is a smaller cavity following the constriction for this phoneme than for the /ʃ/. This causes the noise for the /s/ to appear at higher frequencies than does the noise in /ʃ/ (Strevens, 1960; Heinz & Stevens, 1961; Jassem, 1965; Behrens & Blumstein, 1988). This is apparent in Figure 4, which contains


System
Capture
Data
View
Link
Show
Speak
Analyze
Edit
Tag
Macro
Log
A>ch1 : *Captured* - SCALE
6.00000<
1116>
Amplitude
32K
-32K
0.000
0.972
B>SPG
0.00000<
0>
10156
0
Freq. (Hz)
0.000
0.972
Time (sec)

waveforms of tokens of /sæ/ and /ʃæ/ in the top portion of the figure, and spectrograms in the bottom portion. In the spectrogram, time is on the x-axis, and frequency is on the y-axis. Amplitude is shown by the darkness of the ink in the picture. Thus, the darker sections represent frequencies at which there is more energy. The noise portion at the start of each syllable is the frication. This frication is at higher frequencies in the /s/ than the /ʃ/. Harris (1958; see also Heinz & Stevens, 1961) found that the noise center frequency information is the primary cue for distinguishing these particular phonemes. However, Tomiak's (1991) results suggest that there may be some additional cue listeners' make use of during perception. That is, while the spectral information in the noise is the main cue, it may not be the only cue listeners actually use.

Work on other fricatives, such as /f/ and /θ/, (Harris, 1958) suggests that the formant frequencies at the onset of vocal pulsing may also cue the place-of-articulation distinction between fricatives. During production of voiceless fricatives, the speaker forces air through a small constriction, causing turbulence. At the end of this frication, the vocal folds start vibrating and the articulators move into position for the following vowel.[13]

---

[13] Using the word "vibrate" for the vocal fold action is actually somewhat inaccurate, as it gives the (incorrect) impression that vocal folds act similarly to guitar strings -- that is, that the vibration is the sound source of the noise, much as the vibration of a guitar string is the source of the musical note. Actually, the vocal folds work more in the manner of an old-fashioned airhorn. When the air pressure behind the closed vocal folds is sufficiently high, the vocal folds are blown apart. This causes a puff of air to be released into the supralaryngeal cavity (or the portion of the vocal tract that is "above" the larynx), and

When the vocal folds vibrate, they produce energy at many different frequencies. This energy is then filtered by the vocal tract. Just as with the frication discussed earlier, this filtering emphasizes certain frequencies, and de-emphasizes other frequencies. Changing the shape of the vocal tract changes which frequencies get emphasized. This is physically the same principle that causes different shaped tubes in a pipe organ to produce different sounds -- the length and width of the vocal tract "tube" works in the same way as the length and width of the pipe organ tubes. When the tongue, jaw, and lips move, they change the shape of the "tube", causing different frequencies to be emphasized.

Frequencies that are emphasized appear as dark (roughly horizontal) bands in the spectrogram in Figure 4, and the center frequencies of these formants have been added as a dark line. (Remember, darkness represents amplitude here. So, darker portions are frequencies at which there is more energy.) These bands are known as formants. The first formant is the band of energy with the lowest frequency. The second formant is the next such band, etc. The center frequencies for these bands are different for different tongue and mouth configurations. Thus, the frequencies of the

---

the air pressure behind the vocal folds drops. At this point, there is no longer the pressure forcing the vocal folds to remain open. The tension in the vocal folds themselves (assisted by the drop in air pressure at the edge of the vocal folds caused by the rushing air, otherwise known as the Bernoulli effect) causes them to snap back together. This process happens repeatedly, and these puffs of air are source of vocal tract sound.

formants are related to the position of the articulators in the mouth, and can be a cue to what sound the speaker was trying to produce.

As the speaker begins vocal fold vibration, he moves his articulators away from the positions they held during the fricative and towards the position necessary for the following sound. This causes the formants to change in frequency. During this transition, the formants are indicative of both the position of the articulators during the fricative production, and those necessary for the following phoneme. Thus, the formant values at the onset of voicing (or the offset of frication) can be a cue to the intended fricative.

Formant frequencies at fricative offset (and vowel onset) appear to be a primary cue for distinguishing the fricatives /f/ and /θ/ ("th," as in "thin"; Harris, 1958), but they might be used secondarily for the other fricatives as well. This is supported by research by Whalen (1981) suggesting that the same frication noise can be heard as /s/ before /s/-transitions and as /ʃ/ before /ʃ/-transitions. Also, Whalen (1991), Repp (1981; Mann & Repp, 1980), and Hedrick and Ohde (1993) have found that both the noise spectrum and the transitions into the vowel are typically used in making /s/-/ʃ/ judgments, at least for ambiguous stimuli. Whalen (1984) found that information in the transitions could affect perception of even clear fricative tokens. He cross-spliced /sV/ and /ʃV/ syllables, such that the information in the transition could be appropriate for the consonant (that is, the vowel could have been produced in the same consonantal context) or be inappropriate (that is, the vowel had been originally produced with a different consonant). Even though the frication portions of the syllables were clear cases, this mismatch information in the transitions slowed identification.

The suggestion that formant transitions may be important is also supported by the linguistic notion of an abstract place of articulation. The /s/ is an alveolar fricative, produced in the same place of articulation as are the stop consonants /d/ and /t/. (That is, they are produced with the tongue against the alveolar ridge of the mouth, immediately behind the teeth.) The /ʃ/, on the other hand, is produced with the tongue obstructing the airway near the hard palate, and is thus considered to have a palatal place of articulation. The stop consonants /k/ and /g/ are generally considered velar consonants, but actually have a palatal place of articulation before front vowels (see Ladefoged, 1982; MacKay, 1987). If place of articulation is really an abstract feature of phonemes, we might expect the formant frequencies of /s/ at the start of vocal pulsing to be more like those of /d/ and /t/, and those of /ʃ/ to be more like those of /k/ and /g/. However, as this has been described as being a secondary cue, it is possible that not all speakers would actually use this distinction in production or use it to the same extent.

In order to investigate this issue, I examined a database of 6 different talkers (3 male, 3 female), producing /s/ and /ʃ/ before the vowels /ɑ/ and /æ/. This database was created for another purpose (see Tomiak, 1991), but consisted of individuals speaking a variety of fricative-vowel syllables in isolation. The recordings were made at a 20 kHz sampling rate with an Electro-Voice D054 Dynamic Omni Microphone, and were spoken in the carrier phrase "Please say ____ to me". They were low-pass filtered at 9.6 kHz, and recorded in 12-bit digital format on a DEC VAX 11/730 computer. There were two tokens from each talker, resulting in a total of 48 utterances (6 talkers x 2 consonants x 2 vowels x 2 tokens). For each utterance, the experimenter listened to the utterance with all but the final 15 ms of frication removed, and identified the token. For four of the six talkers, almost all utterances sounded as if they began with a /d/ or /t/. However, for two of the talkers, the tokens that had begun with an /s/ sounded as if they began with a /d/, but those beginning with an /ʃ/ actually sounded as if they began with a /g/. Although this piloting was based on only six talkers, and only one listener, it suggests that formant frequencies might actually be used by some talkers during /s/ and /ʃ/ production.

We now appear to have two potential cues: frequency center (or centroid) of frication, and formant frequencies at vocal onset. In order to examine the perception-production relations for fricatives, two perceptual series were made. One series varied in frequency centroid, the other varied in formant frequencies at vocal onset. This allowed for determining two "prototypes": One prototype for fricative values, and the other for formant values. Participants' productions were measured with respect to both cues, and the values of the two perception-production correlations compared. The first correlation was between the formant values in production and the formant values in the perceptual prototypes, and the second correlation was between frication centroid values in production and the fricative centroid of the perceptual prototype.

Unfortunately, while there is an obvious way of measuring frequency centroids, the formant frequencies are not so easily described by one single value. Both F2 and F3 are important for the /d/-/g/ distinction, and might be expected to be relevant here, as well.[14] Furthermore, both cues seem to differ between productions of /s/ and /ʃ/ (Mann & Repp,

---

[14] A study by Datscheweit (1990) is of relevance here. He examined the influence of F2 onset frequencies on the perception of /s/ and /ʃ/. He found that F2 did have an influence on goodness ratings, but did not serve to differentiate /s/ and /ʃ/. However, he used relatively large step sizes in his alterations of F2, and thus it is possible that small differences may have been missed. Furthermore, although he was intending to examine F2, he kept F3 constant, and thus varied the difference between the formants as well as the F2 formant itself. As more recent research has suggested that the differences between formants may be a more relevant cue than absolute values (Syrdal & Gopal, 1986), we have chosen to examine these difference scores, even though there is a precedent in the literature for examining F2 alone.

1980). Unfortunately, the methodology used here requires that there be one acoustic measure which can be used to evaluate individuals' productions, and which can be varied across stimuli in the perception task. Some literature (Sawusch & Dutton, 1992) suggests that the difference between F3 and F2 might be a reasonable acoustic correlate for the formant difference between /d/ and /g/. In /g/ and /k/, these two formants tend to be much closer to one another at the beginning of the formant transitions than they are in /d/ and /t/. This appears to also be the case for /ʃ/ relative to /s/, since /ʃ/ tends to have a higher F2 and lower F3 at onset (Mann & Repp, 1980). I proceeded to examine whether this carries over to fricatives by measuring F3 - F2 for the 48 utterances described above.

Linear Predictive Coding was used to find the formants for each utterance. The LPC was calculated over the beginning of the formant transitions, starting approximately 15 ms before the start of vocal pulsing, and continuing through the first 5 vocal pulses. The window size was kept at 12 ms, and values for F2 and F3 during the first 3 frames were averaged. When the LPC was unable to find a formant for a particular frame, the value from the 4th frame was included in the average, instead. For each subject, the average F3 - F2 for /ʃ/ tokens was smaller than that for /s/ tokens. There were only a few instances where any given /ʃ/ token had an F3 - F2 value that was as large as that found in the /s/ tokens for that subject. This suggests that the F3-F2 difference may indeed be an appropriate way of measuring formant differences between /s/ and /ʃ/, and will be the method used in this experiment.

Method

Subjects. Twenty-four subjects participated in this experiment, which involved 3 one-hour sessions. Subjects received $15 in compensation upon completion of the third session. Because formant differences were considered a secondary cue to the /s/-/ʃ/ distinction, it was not expected that all subjects would make use of this cue. Thus, subjects were not removed from analysis if their ratings did not fall off towards the extremes of the continuum for this set of items. However, subjects whose ratings did not fall off for either of the two continua were removed from analysis. This accounted for 4 subjects, leaving a total of 20 subjects' data in this experiment. Of these 20 subjects, one had also participated in Experiment 1.

Stimuli. For the production task, a female native talker of English (RSN) recorded four tokens of each CV syllable beginning with either /s/ or /ʃ/ and followed by the 7 vowels /i, e, æ, u, o, ɑ, ʌ/. All of the tokens were amplified, low-pass filtered at 9.5 kHz, digitized via a 16-bit, analog-to-digital converter at a 20 kHz sampling rate and stored on computer disk.

For the perception task, the stimuli consisted of two series ranging from /sæ/ to /ʃæ/ and beyond /ʃæ/. The vowel /æ/ was chosen because it does not entail lip-rounding or protrusion, which can alter the spectral information in the fricative (Soli, 1981), and because it does not contain extreme frequency values that might restrict the movement of formant values at fricative offset. The series were created synthetically, as there is no way to edit a natural continua based on slight formant frequency differences. In order to make items varying on both dimensions of interest (centroid of frication and formant frequencies at onset of vocal pulsing), it was necessary to model a speaker who makes both distinctions in his or her production. For this reason, the stimuli were based on the productions of the speaker in the 6-talker database described above who most clearly made the formant frequency distinctions between /s/ and /ʃ/ (KJR). His voice is also one which is readily mimicked by our speech synthesis program, and his recordings were 100% correctly identified by the listeners in Tomiak's (1991) experiments.

The speaker produced tokens of /sæ/ and /ʃæ/ in the context of the carrier phrase, "Please say ____ to me." As I wanted listeners to attend only to the fricative part of the utterance, not to the vowel, it was important to create a series in which the vowel information was constant. The vowels were not entirely identical in the two base syllables of /sæ/ and

/ʃæ/, so creating a series based on these would have made the vowel portions differ slightly across the series as well. In order to keep vowel information constant across the series, it was necessary to create base syllables in which the vowel information was identical. To do this, the vocalic portion of one production was cross-spliced onto the end of the consonant in the other production (creating /s/ and /ʃ/ tokens that had identical vowel information). However, in order to keep the transitions into the vowel distinct in the two productions, the cross-splicing needed to occur after the onset of the vowel, at least for the items varying in formant transitions. To cross-splice at this location without audible distortions or apparent talker changes required locating two productions (one /s/, one /ʃ/) which had similar fundamental frequencies. Two tokens of KJR's productions were found that met these criteria. Two continua were made on the basis of these items, as described below. One continuum consisted of differences in formant frequencies at the end of the frication, the other consisted of differences in the centroid of frication. Although it would have been optimal to present subjects with items consisting of all possible pairings of these different values, this would have resulted in too many stimuli in the perception portion of the experiment. Thus, only items falling along the two axes of the potential 2-dimensional space were presented in the perceptual task.

For the series varying in frication, the transition and vowel portions of the /s/ and /ʃ/ syllables were removed, leaving only the frication portion of the syllables. This frication portion was 215 ms long. These syllables were then synthesized using the parallel mode of a cascade/parallel synthesizer (Klatt, 1980). Formant[15] frequencies and bandwidths were adjusted to make the synthetic tokens both sound as similar to the original items as possible and look as similar as possible in a spectral cross-section. The vowel portion from one of the syllables was likewise synthesized and its formant values, bandwidths and amplitudes adjusted. This vowel portion was then appended to both the /s/ and /ʃ/ tokens, so that the two endpoints had identical synthesis parameters after the first 215 ms (or 43 frames). Values for the initial frication portion were then interpolated between the two endpoints to make a 21-item series (including the /s/ and /ʃ/ endpoints).

Rather than make the series continue beyond /ʃ/ in an acoustic manner (by continuing to adjust formant and amplitude values in the same manner as in the first half of the continuum), the series continued beyond /ʃ/ in an articulatory sense. That is, rather than adjust the formants to the

---

[15] The term "formant" is used by the Klatt synthesizer to refer to the resonances in the synthesizer. This term is used even when the sound source is noise. So even though we generally refer to noise in the spectra as frication, rather than formants, we still use the term "formants" when creating the stimuli. Altering the frequency of the formants with a fricative sound source is what allows us to change the centroid of frication.

same degree and in the same direction as between /s/ and /ʃ/, formants were adjusted so as to mimic a more extreme place of articulation. A linguist was brought into the laboratory and asked to produce fricatives from a variety of places of articulation: alveolar (as in /s/), palatal (/ʃ/), and velar and uvular fricatives (which do not occur in English but do occur in other languages). The movements of formants between her tokens were analyzed, and the formants in our synthetic continua were adjusted to move in the same manner. That is, our formant movements beyond /ʃ/ were such that they moved towards a more velar place of articulation. A 20-item series was created in this manner, resulting in a total of 41 stimuli (the /s/ endpoint, 19 interpolated items between /s/ and /ʃ/, the /ʃ/ endpoint, 19 interpolated items beyond /ʃ/, and the more velar endpoint, here labeled /*ʃ/). The synthesis parameters for these three endpoint items are shown in Tables 5-7.

For the series varying in formants, the vowel from one of KJR's tokens was cross-spliced onto the other token, so as to make syllables with equivalent vowel information. The frication portion was removed, so as to allow the formant transitions to be altered without changing the frication information. After frication removal, these partial-syllables only differed in their initial 40 ms. These syllables were resynthesized using the parallel mode of the synthesizer, which allowed full control of the amplitude levels

Table 5 /s/ ("s") endpoint, series varying in frication

Global Parameters:

F Glt Res	B Glt Res	F Glt Zero	B Glt Zero	B Glt Res2
0	100	1500	6000	200
F6	**B6**	**F Nsl Pol**	**B Nsl Pol**	**B Nsl Zero**
4900	1000	250	100	100
Gain	**Auto Amp**	**No.Cas For**	**C/P SW**	**Cor SW**
26	-1	5	0	1

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
0	468	41	39	1802	182	27	2700	220	25	3509	230	30	4581	221	57	48	0	250	160	0	0	0	0	65
5	468	41	39	1798	182	27	2700	220	25	3517	230	30	4582	221	57	48	0	250	160	0	0	0	0	66
10	468	41	39	1795	182	27	2700	220	25	3525	230	30	4584	221	57	48	0	250	160	0	0	0	0	66
15	468	41	39	1791	182	27	2700	220	25	3533	230	30	4585	221	57	48	0	250	160	0	0	0	0	67
20	468	41	39	1787	181	27	2700	220	25	3541	230	30	4587	221	57	48	0	250	160	0	0	0	0	67
25	468	41	39	1784	181	27	2700	220	25	3548	230	30	4588	221	57	48	0	250	160	0	0	0	0	68
30	468	41	39	1780	181	27	2700	220	25	3556	230	30	4589	221	57	48	0	250	160	0	0	0	0	68
35	468	41	39	1777	181	27	2700	220	25	3564	230	30	4591	221	57	48	0	250	160	0	0	0	0	69
40	468	41	39	1773	181	27	2700	220	25	3572	230	30	4592	221	57	48	0	250	160	0	0	0	0	69
45	468	41	39	1769	181	27	2700	220	25	3580	230	30	4594	221	57	48	0	250	160	0	0	0	0	70
50	468	41	39	1766	181	27	2700	220	25	3588	230	30	4595	221	57	48	0	250	160	0	0	0	0	70
55	468	41	39	1762	180	27	2700	220	25	3596	230	30	4596	221	57	48	0	250	160	0	0	0	0	71
60	468	41	39	1758	180	27	2700	220	25	3604	230	30	4598	221	57	48	0	250	160	0	0	0	0	71
65	468	41	39	1755	180	27	2700	220	25	3612	230	30	4599	221	57	48	0	250	160	0	0	0	0	72
70	468	41	39	1751	180	27	2700	220	25	3619	230	30	4601	221	57	48	0	250	160	0	0	0	0	72
75	468	41	39	1748	180	27	2700	220	25	3627	230	30	4602	221	57	48	0	250	160	0	0	0	0	73
80	468	41	39	1744	180	27	2700	220	25	3635	230	30	4603	221	57	48	0	250	160	0	0	0	0	73
85	468	41	39	1740	180	27	2700	220	25	3643	230	30	4605	221	57	48	0	250	160	0	0	0	0	73
90	468	41	39	1737	179	27	2700	220	25	3651	230	31	4606	221	56	48	0	250	160	0	0	0	0	73
95	468	41	39	1733	179	27	2700	220	25	3659	230	31	4608	221	56	48	0	250	160	0	0	0	0	74
100	468	41	39	1729	179	27	2700	220	25	3667	230	31	4609	221	56	48	0	250	160	0	0	0	0	74
105	468	41	39	1740	179	27	2700	220	25	3675	230	31	4610	221	56	48	0	250	160	0	0	0	0	74
110	468	41	39	1752	179	27	2700	220	25	3682	230	31	4612	221	56	48	0	250	160	0	0	0	0	74
115	468	41	39	1763	179	27	2700	220	25	3690	230	31	4613	221	56	48	0	250	160	0	0	0	0	74

Table 5, continued /s/ ("s") endpoint, series varying in frication

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
120	468	41	39	1775	179	27	2700	220	25	3698	230	31	4615	221	56	48	0	250	160	0	0	0	0	74
125	468	41	39	1786	178	27	2700	220	25	3706	230	31	4616	221	56	48	0	250	160	0	0	0	0	74
130	468	41	39	1797	178	27	2700	220	25	3714	230	31	4617	234	56	48	0	250	160	0	0	0	0	74
135	468	41	39	1809	178	27	2700	220	25	3722	230	31	4619	247	56	48	0	250	160	0	0	0	0	75
140	468	41	39	1820	178	27	2700	220	25	3730	230	31	4620	260	56	48	0	250	160	0	0	0	0	75
145	468	41	39	1832	178	27	2700	220	25	3738	230	31	4622	273	56	48	0	250	160	0	0	0	0	75
150	468	41	39	1843	178	27	2700	220	25	3746	230	31	4623	285	56	48	0	250	160	0	0	0	0	73
155	468	41	39	1854	178	27	2700	220	25	3743	230	31	4624	298	56	48	0	250	160	0	0	0	0	70
160	468	41	39	1866	177	27	2700	220	25	3739	230	31	4626	311	56	48	0	250	160	0	0	0	0	68
165	468	41	39	1877	177	27	2700	220	25	3736	230	31	4627	324	56	48	0	250	160	0	0	0	0	65
170	468	41	39	1889	177	27	2700	220	25	3733	230	31	4629	337	56	48	0	250	160	0	0	0	0	63
175	468	41	39	1900	177	27	2700	220	25	3729	230	31	4630	350	56	48	0	250	160	0	0	0	0	60
180	468	41	62	1892	173	48	2686	228	42	3726	230	38	4632	350	31	0	0	250	160	60	0	0	0	0
185	468	41	59	1885	170	48	2672	236	42	3723	248	38	4607	363	30	0	0	250	159	67	0	0	0	0
190	468	41	55	1877	166	48	2658	244	42	3720	265	38	4582	375	28	0	0	250	158	70	0	0	0	0
195	484	41	52	1870	162	48	2644	252	42	3716	283	36	4557	388	27	0	0	250	157	73	0	0	0	0
200	488	40	49	1862	159	68	2630	260	42	3713	296	34	4532	400	26	0	0	250	156	73	0	0	0	0
205	489	40	47	1854	155	59	2639	267	31	3698	310	31	4507	413	20	0	0	250	155	74	0	0	0	0
210	503	40	44	1847	125	60	2648	264	32	3672	323	29	4482	475	20	0	0	250	155	74	0	0	0	0
215	530	40	41	1839	107	60	2657	247	32	3652	337	27	4457	488	20	0	0	250	154	74	0	0	0	0
220	559	40	41	1797	100	61	2666	228	33	3631	350	25	4432	500	20	0	0	250	154	76	0	0	0	0
225	575	39	43	1779	96	61	2676	209	33	3608	350	24	4432	500	20	0	0	250	153	77	0	0	0	0
230	586	38	44	1762	94	61	2677	203	33	3628	350	24	4432	500	20	0	0	250	153	76	0	0	0	0
235	598	37	45	1765	91	62	2669	202	33	3648	350	23	4432	500	20	0	0	250	153	76	0	0	0	0
240	606	37	45	1756	84	62	2651	211	32	3669	350	20	4432	500	20	0	0	250	153	77	0	0	0	0
245	612	36	44	1746	73	63	2633	226	31	3689	350	19	4432	500	20	0	0	250	152	76	0	0	0	0
250	619	35	44	1739	64	64	2615	226	30	3709	350	20	4432	500	20	0	0	250	152	77	0	0	0	0
255	627	35	44	1738	64	65	2597	226	31	3729	350	19	4432	500	20	0	0	250	152	77	0	0	0	0
260	638	34	43	1740	64	66	2579	226	30	3749	350	19	4432	500	20	0	0	250	152	77	0	0	0	0
265	649	37	43	1735	64	67	2561	226	30	3769	350	19	4432	500	20	0	0	250	151	77	0	0	0	0
270	663	37	43	1727	64	66	2543	227	30	3790	350	17	4432	500	20	0	0	250	151	77	0	0	0	0
275	672	36	43	1724	63	66	2518	227	30	3810	350	14	4432	500	20	0	0	250	150	77	0	0	0	0
280	679	36	42	1717	63	66	2492	227	29	3830	350	15	4432	500	20	0	0	250	150	76	0	0	0	0
285	685	39	41	1717	63	66	2494	227	30	3837	350	14	4432	500	20	0	0	250	149	75	0	0	0	0
290	690	40	42	1710	63	66	2510	227	31	3844	350	16	4432	500	20	0	0	250	148	75	0	0	0	0
295	696	41	41	1703	63	66	2512	219	31	3851	350	17	4432	500	20	0	0	250	147	75	0	0	0	0

Table 5, continued /s/ ("s") endpoint, series varying in frication

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
300	698	41	40	1699	62	66	2509	216	30	3851	350	15	4432	500	20	0	0	250	147	74	0	0	0	0
305	707	42	38	1690	60	65	2491	213	29	3851	350	17	4432	500	20	0	0	250	146	74	0	0	0	0
310	716	43	38	1683	60	67	2496	210	30	3851	350	20	4432	500	20	0	0	250	146	75	0	0	0	0
315	722	43	38	1679	58	66	2505	207	29	3852	350	19	4432	500	20	0	0	250	145	73	0	0	0	0
320	733	44	38	1675	60	66	2515	204	29	3852	350	19	4432	500	20	0	0	250	145	71	0	0	0	0
325	739	44	38	1666	61	65	2522	200	28	3852	350	17	4432	500	20	0	0	250	144	71	0	0	0	0
330	751	44	39	1659	67	65	2509	197	29	3852	350	18	4432	500	20	0	0	250	144	70	0	0	0	0
335	768	44	39	1668	72	64	2475	194	29	3852	350	19	4432	500	20	0	0	250	143	70	0	0	0	0
340	781	44	38	1679	83	61	2494	191	27	3852	350	17	4432	500	20	0	0	250	143	71	0	0	0	0
345	780	44	40	1681	84	61	2499	188	28	3853	350	16	4432	500	20	0	0	250	142	68	0	0	0	0
350	781	44	39	1677	85	59	2494	185	27	3853	350	17	4432	500	20	0	0	250	141	70	0	0	0	0
355	777	44	39	1673	86	60	2492	179	28	3853	350	13	4432	500	20	0	0	250	140	70	0	0	0	0
360	773	42	39	1681	87	59	2493	191	26	3853	350	9	4432	500	20	0	0	250	139	69	0	0	0	0
365	772	42	40	1684	88	58	2507	193	26	3853	350	12	4432	500	20	0	0	250	138	69	0	0	0	0
370	772	42	40	1656	89	58	2516	193	26	3853	350	9	4432	500	20	0	0	250	137	69	0	0	0	0
375	772	43	39	1648	90	57	2527	172	25	3853	350	9	4432	500	20	0	0	250	136	69	0	0	0	0
380	767	43	40	1656	91	57	2530	165	26	3854	350	13	4432	500	20	0	0	250	135	69	0	0	0	0
385	754	43	38	1668	92	58	2504	158	24	3854	350	12	4432	500	20	0	0	250	134	68	0	0	0	0
390	743	43	35	1661	82	57	2496	164	23	3854	350	10	4432	500	20	0	0	250	133	68	0	0	0	0
395	730	43	36	1646	75	57	2507	160	24	3854	350	11	4432	500	20	0	0	250	132	67	0	0	0	0
400	713	44	34	1648	71	55	2533	170	21	3854	350	13	4432	500	20	0	0	250	131	67	0	0	0	0
405	707	44	32	1646	67	52	2525	201	18	3854	350	10	4432	500	20	0	0	250	130	66	0	0	0	0
410	696	44	31	1639	76	49	2496	211	15	3855	350	9	4432	500	20	0	0	250	129	64	0	0	0	0
415	697	44	27	1668	85	43	2489	208	14	3855	350	8	4432	500	20	0	0	250	128	63	0	0	0	0
420	688	44	23	1660	95	40	2504	181	12	3855	350	5	4432	500	20	0	0	250	128	61	0	0	0	0
425	666	45	21	1663	84	38	2504	230	12	3855	350	5	4432	500	20	0	0	250	127	55	0	0	0	0
430	677	69	12	1663	86	33	2492	283	-5	3835	350	-3	4432	500	20	0	0	250	127	48	0	0	0	0
435	663	81	3	1665	80	23	2492	336	-5	3835	350	-3	4432	500	20	0	0	250	125	42	0	0	0	0

Table 6 /ʃ/ ("sh") endpoint, series varying in frication

Global Parameters:

F Glt Res	B Glt Res	F Glt Zero	B Glt Zero	B Glt Res2
0	100	1500	6000	200
F6	B6	F Nsl Pol	B Nsl Pol	B Nsl Zero
4900	1000	250	100	100
Gain	Auto Amp	No.Cas For	C/P SW	Cor SW
26	-1	5	0	1

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
0	468	130	65	2050	340	48	2700	220	45	3670	230	39	4220	350	37	0	0	250	160	0	0	0	0	65
5	468	130	65	2046	337	48	2700	220	45	3670	230	39	4232	350	37	0	0	250	160	0	0	0	0	66
10	468	130	65	2041	333	48	2700	220	45	3670	230	39	4243	350	37	0	0	250	160	0	0	0	0	66
15	468	130	65	2037	330	48	2700	220	45	3670	230	39	4255	350	37	0	0	250	160	0	0	0	0	67
20	468	130	65	2033	327	48	2700	220	45	3670	230	39	4267	350	37	0	0	250	160	0	0	0	0	67
25	468	130	65	2029	324	48	2700	220	45	3670	230	39	4279	350	37	0	0	250	160	0	0	0	0	68
30	468	130	65	2024	320	48	2700	220	45	3670	230	39	4290	350	37	0	0	250	160	0	0	0	0	68
35	468	130	65	2020	317	48	2700	220	45	3670	230	39	4302	350	37	0	0	250	160	0	0	0	0	69
40	468	130	65	2016	314	48	2700	220	45	3670	230	39	4314	350	37	0	0	250	160	0	0	0	0	69
45	468	130	65	2011	310	48	2700	220	45	3670	230	40	4325	350	36	0	0	250	160	0	0	0	0	70
50	468	130	65	2007	307	48	2700	220	45	3670	230	40	4337	350	36	0	0	250	160	0	0	0	0	70
55	468	130	65	2003	304	48	2700	220	45	3670	230	40	4349	350	36	0	0	250	160	0	0	0	0	71
60	468	130	65	1999	301	48	2700	220	45	3670	230	40	4361	350	36	0	0	250	160	0	0	0	0	71
65	468	130	65	1994	297	48	2700	220	45	3670	230	40	4372	350	36	0	0	250	160	0	0	0	0	72
70	468	130	65	1990	294	48	2700	220	45	3670	230	40	4384	350	36	0	0	250	160	0	0	0	0	72
75	468	130	65	1986	291	48	2700	220	45	3670	230	40	4396	350	36	0	0	250	160	0	0	0	0	73
80	468	130	65	1981	287	48	2700	220	45	3670	230	40	4407	350	36	0	0	250	160	0	0	0	0	73
85	468	130	65	1977	284	48	2700	220	45	3670	230	40	4419	350	36	0	0	250	160	0	0	0	0	73
90	468	130	65	1973	281	48	2700	220	45	3670	230	40	4431	350	36	0	0	250	160	0	0	0	0	73
95	468	130	65	1969	278	48	2700	220	45	3670	230	40	4443	350	36	0	0	250	160	0	0	0	0	74
100	468	130	65	1964	274	48	2700	220	45	3670	230	40	4454	350	36	0	0	250	160	0	0	0	0	74
105	468	130	65	1960	271	48	2700	220	45	3670	230	40	4466	350	36	0	0	250	160	0	0	0	0	74
110	468	130	65	1956	268	48	2700	220	45	3670	230	40	4478	350	36	0	0	250	160	0	0	0	0	74
115	468	130	65	1951	264	48	2700	220	45	3670	230	40	4489	350	36	0	0	250	160	0	0	0	0	74

Table 6, continued /ʃ/ ("sh") endpoint, series varying in frication

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
120	468	130	65	1947	261	48	2700	220	45	3670	230	40	4501	350	36	0	0	250	160	0	0	0	0	74
125	468	130	65	1943	258	48	2700	220	45	3670	230	40	4513	350	36	0	0	250	160	0	0	0	0	74
130	468	130	65	1939	255	48	2700	220	45	3670	230	40	4525	350	36	0	0	250	160	0	0	0	0	74
135	468	130	65	1934	251	48	2700	220	45	3670	230	41	4536	350	35	0	0	250	160	0	0	0	0	75
140	468	130	65	1930	248	48	2700	220	45	3670	230	41	4548	350	35	0	0	250	160	0	0	0	0	75
145	468	130	65	1926	245	48	2700	220	45	3670	230	41	4560	350	35	0	0	250	160	0	0	0	0	75
150	468	130	65	1921	241	48	2700	220	45	3670	230	41	4571	350	35	0	0	250	160	0	0	0	0	73
155	468	130	65	1917	238	48	2700	220	45	3670	230	41	4583	350	35	0	0	250	160	0	0	0	0	70
160	468	130	65	1913	235	48	2700	220	45	3670	230	41	4595	350	35	0	0	250	160	0	0	0	0	68
165	468	130	65	1909	232	48	2700	220	45	3670	230	41	4607	350	35	0	0	250	160	0	0	0	0	65
170	468	130	65	1904	228	48	2700	220	45	3670	230	41	4618	350	35	0	0	250	160	0	0	0	0	63
175	468	130	65	1900	225	48	2700	220	45	3670	230	41	4630	350	35	0	0	250	160	0	0	0	0	60
180	468	130	62	1892	219	48	2686	228	42	3686	230	38	4632	350	31	0	0	250	160	60	0	0	0	0
185	468	130	59	1885	213	48	2672	236	42	3680	248	38	4607	363	30	0	0	250	159	67	0	0	0	0
190	468	130	55	1877	207	48	2658	244	42	3682	265	38	4582	375	28	0	0	250	158	70	0	0	0	0
195	484	130	52	1870	201	48	2644	252	42	3690	283	36	4557	388	27	0	0	250	157	73	0	0	0	0
200	488	97	49	1862	195	68	2630	260	42	3713	296	34	4532	400	26	0	0	250	156	73	0	0	0	0
205	489	60	47	1854	155	59	2639	267	31	3698	310	31	4507	413	20	0	0	250	155	74	0	0	0	0
210	503	41	44	1847	125	60	2648	264	32	3672	323	29	4482	475	20	0	0	250	155	74	0	0	0	0
215	530	40	41	1839	107	60	2657	247	32	3652	337	27	4457	488	20	0	0	250	154	74	0	0	0	0
220	559	40	41	1797	100	61	2666	228	33	3631	350	25	4432	500	20	0	0	250	154	76	0	0	0	0
225	575	39	43	1779	96	61	2676	209	33	3608	350	24	4432	500	20	0	0	250	153	77	0	0	0	0
230	586	38	44	1762	94	61	2677	203	33	3628	350	24	4432	500	20	0	0	250	153	76	0	0	0	0
235	598	37	45	1765	91	62	2669	202	33	3648	350	23	4432	500	20	0	0	250	153	76	0	0	0	0
240	606	37	45	1756	84	62	2651	211	32	3669	350	20	4432	500	20	0	0	250	153	77	0	0	0	0
245	612	36	44	1746	73	63	2633	226	31	3689	350	19	4432	500	20	0	0	250	152	76	0	0	0	0
250	619	35	44	1739	64	64	2615	226	30	3709	350	20	4432	500	20	0	0	250	152	77	0	0	0	0
255	627	35	44	1738	64	65	2597	226	31	3729	350	19	4432	500	20	0	0	250	152	77	0	0	0	0
260	638	34	43	1740	64	66	2579	226	30	3749	350	19	4432	500	20	0	0	250	152	77	0	0	0	0
265	649	37	43	1735	64	67	2561	226	30	3769	350	19	4432	500	20	0	0	250	151	77	0	0	0	0
270	663	37	43	1727	64	66	2543	227	30	3790	350	17	4432	500	20	0	0	250	151	77	0	0	0	0
275	672	36	43	1724	63	66	2518	227	30	3810	350	14	4432	500	20	0	0	250	150	77	0	0	0	0
280	679	36	42	1717	63	66	2492	227	29	3830	350	15	4432	500	20	0	0	250	150	76	0	0	0	0
285	685	39	41	1717	63	66	2494	227	30	3837	350	14	4432	500	20	0	0	250	149	75	0	0	0	0
290	690	40	42	1710	63	66	2510	227	31	3844	350	16	4432	500	20	0	0	250	148	75	0	0	0	0
295	696	41	41	1703	63	66	2512	219	31	3851	350	17	4432	500	20	0	0	250	147	75	0	0	0	0

Table 6, continued /ʃ/ ("sh") endpoint, series varying in frication

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
300	698	41	40	1699	62	66	2509	216	30	3851	350	15	4432	500	20	0	0	250	147	74	0	0	0	0
305	707	42	38	1690	60	65	2491	213	29	3851	350	17	4432	500	20	0	0	250	146	74	0	0	0	0
310	716	43	38	1683	60	67	2496	210	30	3851	350	20	4432	500	20	0	0	250	146	75	0	0	0	0
315	722	43	38	1679	58	66	2505	207	29	3852	350	19	4432	500	20	0	0	250	145	73	0	0	0	0
320	733	44	38	1675	60	66	2515	204	29	3852	350	19	4432	500	20	0	0	250	145	71	0	0	0	0
325	739	44	38	1666	61	65	2522	200	28	3852	350	17	4432	500	20	0	0	250	144	71	0	0	0	0
330	751	44	39	1659	67	65	2509	197	29	3852	350	18	4432	500	20	0	0	250	144	70	0	0	0	0
335	768	44	39	1668	72	64	2475	194	29	3852	350	19	4432	500	20	0	0	250	143	70	0	0	0	0
340	781	44	38	1679	83	61	2494	191	27	3852	350	17	4432	500	20	0	0	250	143	71	0	0	0	0
345	780	44	40	1681	84	61	2499	188	28	3853	350	16	4432	500	20	0	0	250	142	68	0	0	0	0
350	781	44	39	1677	85	59	2494	185	27	3853	350	17	4432	500	20	0	0	250	141	70	0	0	0	0
355	777	44	39	1673	86	60	2492	179	28	3853	350	13	4432	500	20	0	0	250	140	70	0	0	0	0
360	773	42	39	1681	87	59	2493	191	26	3853	350	9	4432	500	20	0	0	250	139	69	0	0	0	0
365	772	42	40	1684	88	58	2507	193	26	3853	350	12	4432	500	20	0	0	250	138	69	0	0	0	0
370	772	42	40	1656	89	58	2516	193	26	3853	350	9	4432	500	20	0	0	250	137	69	0	0	0	0
375	772	43	39	1648	90	57	2527	172	25	3853	350	9	4432	500	20	0	0	250	136	69	0	0	0	0
380	767	43	40	1656	91	57	2530	165	26	3854	350	13	4432	500	20	0	0	250	135	69	0	0	0	0
385	754	43	38	1668	92	58	2504	158	24	3854	350	12	4432	500	20	0	0	250	134	68	0	0	0	0
390	743	43	35	1661	82	57	2496	164	23	3854	350	10	4432	500	20	0	0	250	133	68	0	0	0	0
395	730	43	36	1646	75	57	2507	160	24	3854	350	11	4432	500	20	0	0	250	132	67	0	0	0	0
400	713	44	34	1648	71	55	2533	170	21	3854	350	13	4432	500	20	0	0	250	131	67	0	0	0	0
405	707	44	32	1646	67	52	2525	201	18	3854	350	10	4432	500	20	0	0	250	130	66	0	0	0	0
410	696	44	31	1639	76	49	2496	211	15	3855	350	9	4432	500	20	0	0	250	129	64	0	0	0	0
415	697	44	27	1668	85	43	2489	208	14	3855	350	8	4432	500	20	0	0	250	128	63	0	0	0	0
420	688	44	23	1660	95	40	2504	181	12	3855	350	5	4432	500	20	0	0	250	128	61	0	0	0	0
425	666	45	21	1663	84	38	2504	230	12	3855	350	5	4432	500	20	0	0	250	127	55	0	0	0	0
430	677	69	12	1663	86	33	2492	283	-5	3835	350	-3	4432	500	20	0	0	250	127	48	0	0	0	0
435	663	81	3	1665	80	23	2492	336	-5	3835	350	-3	4432	500	20	0	0	250	125	42	0	0	0	0

Table 7 /*ʃ/ (beyond "sh") endpoint, series varying in frication

Global Parameters:

F Glt Res	B Glt Res	F Glt Zero	B Glt Zero	B Glt Res2
0	100	1500	6000	200
F6	**B6**	**F Nsl Pol**	**B Nsl Pol**	**B Nsl Zero**
4900	1000	250	100	100
Gain	**Auto Amp**	**No.Cas For**	**C/P SW**	**Cor SW**
26	-1	5	0	1

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
0	468	153	72	1713	272	52	2700	400	0	3702	230	0	4130	383	0	0	40	250	160	0	0	0	0	65
5	468	153	72	1709	268	52	2700	400	0	3700	230	0	4144	383	0	0	40	250	160	0	0	0	0	66
10	468	153	72	1704	264	52	2700	400	0	3699	230	0	4159	383	0	0	40	250	160	0	0	0	0	66
15	468	153	72	1700	261	52	2700	400	0	3697	230	0	4173	383	0	0	40	250	160	0	0	0	0	67
20	468	153	72	1695	257	52	2700	400	0	3695	230	0	4187	383	0	0	40	250	160	0	0	0	0	67
25	468	153	72	1691	253	52	2700	400	0	3693	230	0	4201	383	0	0	40	250	160	0	0	0	0	68
30	468	153	72	1686	249	52	2700	400	0	3692	230	1	4216	383	0	0	40	250	160	0	0	0	0	68
35	468	153	72	1682	245	52	2700	400	0	3690	230	1	4230	383	0	0	40	250	160	0	0	0	0	69
40	468	153	72	1677	241	52	2700	400	0	3688	230	1	4244	383	0	0	40	250	160	0	0	0	0	69
45	468	153	72	1673	238	52	2700	400	0	3687	230	1	4259	383	0	0	40	250	160	0	0	0	0	70
50	468	153	72	1668	234	52	2700	400	0	3685	230	1	4273	383	0	0	40	250	160	0	0	0	0	70
55	468	153	72	1664	230	52	2700	400	0	3683	230	1	4287	383	0	0	40	250	160	0	0	0	0	71
60	468	153	72	1659	226	52	2700	400	0	3681	230	1	4301	383	0	0	40	250	160	0	0	0	0	71
65	468	153	72	1655	222	52	2700	400	0	3680	230	1	4316	383	0	0	40	250	160	0	0	0	0	72
70	468	153	72	1650	218	52	2700	400	0	3678	230	1	4330	383	0	0	40	250	160	0	0	0	0	72
75	468	153	72	1646	215	52	2700	400	0	3676	230	1	4344	383	0	0	40	250	160	0	0	0	0	73
80	468	153	72	1641	211	52	2700	400	0	3675	230	1	4359	383	0	0	40	250	160	0	0	0	0	73
85	468	153	72	1637	207	52	2700	400	0	3673	230	1	4373	383	0	0	40	250	160	0	0	0	0	73
90	468	153	72	1632	203	52	2700	400	0	3671	230	2	4387	383	0	0	40	250	160	0	0	0	0	73
95	468	153	72	1628	199	52	2700	400	0	3669	230	2	4401	383	0	0	40	250	160	0	0	0	0	74
100	468	153	72	1623	195	52	2700	400	0	3668	230	2	4416	383	0	0	40	250	160	0	0	0	0	74
105	468	153	72	1615	192	52	2700	400	0	3666	230	2	4430	383	0	0	40	250	160	0	0	0	0	74
110	468	153	72	1607	188	52	2700	400	0	3664	230	2	4444	383	0	0	40	250	160	0	0	0	0	74
115	468	153	72	1598	184	52	2700	400	0	3663	230	2	4459	383	0	0	40	250	160	0	0	0	0	74

Table 7, continued /*ʃ/ (beyond "sh") endpoint, series varying in frication

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
120	468	153	72	1590	180	52	2700	400	0	3661	230	2	4473	383	0	0	40	250	160	0	0	0	0	74
125	468	153	72	1582	176	52	2700	400	0	3659	230	2	4487	383	0	0	40	250	160	0	0	0	0	74
130	468	153	72	1574	172	52	2700	400	0	3657	230	2	4501	380	0	0	40	250	160	0	0	0	0	74
135	468	153	72	1566	169	52	2700	400	0	3656	230	2	4516	376	0	0	40	250	160	0	0	0	0	75
140	468	153	72	1557	165	52	2700	400	0	3654	230	2	4530	373	0	0	40	250	160	0	0	0	0	75
145	468	153	72	1549	161	52	2700	400	0	3652	230	2	4544	370	0	0	40	250	160	0	0	0	0	75
150	468	153	72	1541	157	52	2700	400	0	3651	230	3	4559	367	0	0	40	250	160	0	0	0	0	73
155	468	153	72	1533	153	52	2700	400	0	3657	230	3	4573	363	0	0	40	250	160	0	0	0	0	70
160	468	153	72	1525	149	52	2700	400	0	3663	230	3	4587	360	0	0	40	250	160	0	0	0	0	68
165	468	153	72	1516	146	52	2700	400	0	3670	230	3	4601	357	0	0	40	250	160	0	0	0	0	65
170	468	153	72	1508	142	52	2700	400	0	3676	230	3	4616	353	0	0	40	250	160	0	0	0	0	63
175	468	153	72	1500	138	52	2700	400	0	3682	230	3	4630	350	0	0	40	250	160	0	0	0	0	60
180	468	139	62	1892	224	48	2686	228	42	3688	230	38	4632	350	31	0	0	250	160	60	0	0	0	0
185	468	125	59	1885	210	48	2672	236	42	3694	248	38	4607	363	30	0	0	250	159	67	0	0	0	0
190	468	111	55	1877	197	48	2658	244	42	3701	265	38	4582	375	28	0	0	250	158	70	0	0	0	0
195	484	97	52	1870	183	48	2644	252	42	3707	283	36	4557	388	27	0	0	250	157	73	0	0	0	0
200	488	82	49	1862	169	68	2630	260	42	3713	296	34	4532	400	26	0	0	250	156	73	0	0	0	0
205	489	68	47	1854	155	59	2639	267	31	3698	310	31	4507	413	20	0	0	250	155	74	0	0	0	0
210	503	54	44	1847	125	60	2648	264	32	3672	323	29	4482	475	20	0	0	250	155	74	0	0	0	0
215	530	40	41	1839	107	60	2657	247	32	3652	337	27	4457	488	20	0	0	250	154	74	0	0	0	0
220	559	40	41	1797	100	61	2666	228	33	3631	350	25	4432	500	20	0	0	250	154	76	0	0	0	0
225	575	39	43	1779	96	61	2676	209	33	3608	350	24	4432	500	20	0	0	250	153	77	0	0	0	0
230	586	38	44	1762	94	61	2677	203	33	3628	350	24	4432	500	20	0	0	250	153	76	0	0	0	0
235	598	37	45	1765	91	62	2669	202	33	3648	350	23	4432	500	20	0	0	250	153	76	0	0	0	0
240	606	37	45	1756	84	62	2651	211	32	3669	350	20	4432	500	20	0	0	250	153	77	0	0	0	0
245	612	36	44	1746	73	63	2633	226	31	3689	350	19	4432	500	20	0	0	250	152	76	0	0	0	0
250	619	35	44	1739	64	64	2615	226	30	3709	350	20	4432	500	20	0	0	250	152	77	0	0	0	0
255	627	35	44	1738	64	65	2597	226	31	3729	350	19	4432	500	20	0	0	250	152	77	0	0	0	0
260	638	34	43	1740	64	66	2579	226	30	3749	350	19	4432	500	20	0	0	250	152	77	0	0	0	0
265	649	37	43	1735	64	67	2561	226	30	3769	350	19	4432	500	20	0	0	250	151	77	0	0	0	0
270	663	37	43	1727	64	66	2543	227	30	3790	350	17	4432	500	20	0	0	250	151	77	0	0	0	0
275	672	36	43	1724	63	66	2518	227	30	3810	350	14	4432	500	20	0	0	250	150	77	0	0	0	0
280	679	36	42	1717	63	66	2492	227	29	3830	350	15	4432	500	20	0	0	250	150	76	0	0	0	0
285	685	39	41	1717	63	66	2494	227	30	3837	350	14	4432	500	20	0	0	250	149	75	0	0	0	0
290	690	40	42	1710	63	66	2510	227	31	3844	350	16	4432	500	20	0	0	250	148	75	0	0	0	0
295	696	41	41	1703	63	66	2512	219	31	3851	350	17	4432	500	20	0	0	250	147	75	0	0	0	0

Table 7, continued /*ʃ/ (beyond "sh") endpoint, series varying in frication

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
300	698	41	40	1699	62	66	2509	216	30	3851	350	15	4432	500	20	0	0	250	147	74	0	0	0	0
305	707	42	38	1690	60	65	2491	213	29	3851	350	17	4432	500	20	0	0	250	146	74	0	0	0	0
310	716	43	38	1683	60	67	2496	210	30	3851	350	20	4432	500	20	0	0	250	146	75	0	0	0	0
315	722	43	38	1679	58	66	2505	207	29	3852	350	19	4432	500	20	0	0	250	145	73	0	0	0	0
320	733	44	38	1675	60	66	2515	204	29	3852	350	19	4432	500	20	0	0	250	145	71	0	0	0	0
325	739	44	38	1666	61	65	2522	200	28	3852	350	17	4432	500	20	0	0	250	144	71	0	0	0	0
330	751	44	39	1659	67	65	2509	197	29	3852	350	18	4432	500	20	0	0	250	144	70	0	0	0	0
335	768	44	39	1668	72	64	2475	194	29	3852	350	19	4432	500	20	0	0	250	143	70	0	0	0	0
340	781	44	38	1679	83	61	2494	191	27	3852	350	17	4432	500	20	0	0	250	143	71	0	0	0	0
345	780	44	40	1681	84	61	2499	188	28	3853	350	16	4432	500	20	0	0	250	142	68	0	0	0	0
350	781	44	39	1677	85	59	2494	185	27	3853	350	17	4432	500	20	0	0	250	141	70	0	0	0	0
355	777	44	39	1673	86	60	2492	179	28	3853	350	13	4432	500	20	0	0	250	140	70	0	0	0	0
360	773	42	39	1681	87	59	2493	191	26	3853	350	9	4432	500	20	0	0	250	139	69	0	0	0	0
365	772	42	40	1684	88	58	2507	193	26	3853	350	12	4432	500	20	0	0	250	138	69	0	0	0	0
370	772	42	40	1656	89	58	2516	193	26	3853	350	9	4432	500	20	0	0	250	137	69	0	0	0	0
375	772	43	39	1648	90	57	2527	172	25	3853	350	9	4432	500	20	0	0	250	136	69	0	0	0	0
380	767	43	40	1656	91	57	2530	165	26	3854	350	13	4432	500	20	0	0	250	135	69	0	0	0	0
385	754	43	38	1668	92	58	2504	158	24	3854	350	12	4432	500	20	0	0	250	134	68	0	0	0	0
390	743	43	35	1661	82	57	2496	164	23	3854	350	10	4432	500	20	0	0	250	133	68	0	0	0	0
395	730	43	36	1646	75	57	2507	160	24	3854	350	11	4432	500	20	0	0	250	132	67	0	0	0	0
400	713	44	34	1648	71	55	2533	170	21	3854	350	13	4432	500	20	0	0	250	131	67	0	0	0	0
405	707	44	32	1646	67	52	2525	201	18	3854	350	10	4432	500	20	0	0	250	130	66	0	0	0	0
410	696	44	31	1639	76	49	2496	211	15	3855	350	9	4432	500	20	0	0	250	129	64	0	0	0	0
415	697	44	27	1668	85	43	2489	208	14	3855	350	8	4432	500	20	0	0	250	128	63	0	0	0	0
420	688	44	23	1660	95	40	2504	181	12	3855	350	5	4432	500	20	0	0	250	128	61	0	0	0	0
425	666	45	21	1663	84	38	2504	230	12	3855	350	5	4432	500	20	0	0	250	127	55	0	0	0	0
430	677	69	12	1663	86	33	2492	283	-5	3835	350	-3	4432	500	20	0	0	250	127	48	0	0	0	0
435	663	81	3	1665	80	23	2492	336	-5	3835	350	-3	4432	500	20	0	0	250	125	42	0	0	0	0

for each formant. The formant frequency values for these items were smoothed, and amplitude and bandwidth values altered so as to make the synthetic tokens sound as similar to the original items as possible. This resulted in two endpoint items, representing /s/ and /ʃ/. The values for all 5 formants, bandwidths, and amplitudes were then interpolated between the two endpoints, to make an additional 19 items. These changes were then continued to make 20 syllables beyond the /ʃ/ token, varying in the same manner as the items between the /s/ and /ʃ/ tokens. The 20th item in this series is labeled as /*ʃ/. Synthesis parameters for these endpoint items (the good /s/, /ʃ/ and the /*ʃ/ tokens, or items numbered 1, 21, and 41) are shown in Tables 8, 9, and 10. Once this series was created, a frication portion was appended to the beginning of each syllable. It was necessary to select a frication value that was not so salient as to prevent the varying formant frequencies from changing individuals' perceptions. This value was selected on the basis of pilot testing.

Procedure. The procedure was identical to that used in Experiment 1, except that listeners in the perception task were asked to rate the phonemes as examples of the sound "sh", rather than as examples of the sound "p". The subjects participated in 3 1-hour sessions. At the start of the first session, subjects took part in the production task, which used the same procedure as the production task from Experiment 1. There were a

Table 8 /s/ ("s") endpoint, series varying in formant transitions

Global Parameters:

F Glt Res	B Glt Res	F Glt Zero	B Glt Zero	B Glt Res2
0	100	1500	6000	200
F6	B6	F Nsl Pol	B Nsl Pol	B Nsl Zero
5000	1000	250	100	100
Gain	Auto Amp	No.Cas For	C/P SW	Cor SW
29	-1	5	1	0

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
0	514	40	79	1672	85	55	2755	120	60	3710	200	50	4875	300	40	0	0	250	1	0	0	0	0	65
5	513	45	80	1671	85	62	2755	121	61	3721	200	51	4850	300	10	0	0	250	152	57	0	0	0	0
10	519	50	80	1669	85	69	2753	121	62	3742	201	53	4800	300	35	0	0	250	152	60	0	0	0	0
15	538	54	80	1666	82	74	2750	116	62	3744	203	54	4700	266	60	0	0	250	153	63	0	0	0	0
20	565	58	79	1664	79	78	2746	111	63	3746	206	55	4600	233	60	0	0	250	142	63	0	0	0	0
25	585	58	77	1667	84	77	2742	105	66	3748	228	51	4650	242	60	0	0	250	143	63	0	0	0	0
30	600	58	74	1676	88	75	2741	100	68	3700	250	46	4700	250	60	0	0	250	144	63	0	0	0	0
35	599	58	72	1748	82	71	2734	148	69	3693	222	50	4750	225	60	0	0	250	145	63	0	0	0	0
40	592	59	65	1820	75	66	2733	195	69	3686	193	54	4800	200	60	0	0	250	147	63	0	0	0	0
45	586	60	68	1825	81	71	2730	188	69	3648	186	55	4800	200	60	0	0	250	147	63	0	0	0	0
50	591	57	71	1829	86	76	2731	170	66	3610	177	55	4783	200	60	0	0	250	148	63	0	0	0	0
55	599	55	71	1820	85	75	2725	151	64	3605	185	50	4750	200	60	0	0	250	148	63	0	0	0	0
60	606	57	70	1811	78	74	2710	140	62	3599	226	44	4716	200	60	0	0	250	149	63	0	0	0	0
65	612	60	70	1802	69	71	2685	153	64	3625	263	44	4695	225	60	0	0	250	149	63	0	0	0	0
70	620	64	68	1799	61	67	2658	174	65	3650	288	44	4673	250	60	0	0	250	149	63	0	0	0	0
75	629	55	74	1798	57	70	2643	186	65	3725	281	42	4677	275	60	0	0	250	149	63	0	0	0	0
80	640	45	80	1796	55	72	2623	210	65	3800	275	42	4681	300	60	0	0	250	150	63	0	0	0	0
85	655	41	80	1792	60	70	2605	211	66	3767	268	42	4685	300	60	0	0	250	149	59	0	0	0	0
90	670	37	79	1787	65	67	2578	212	66	3733	261	42	4689	300	60	0	0	250	149	60	0	0	0	0
95	675	44	77	1782	61	67	2562	212	64	3817	259	42	4719	300	60	0	0	250	149	61	0	0	0	0
100	681	50	75	1778	57	66	2551	213	62	3900	258	42	4750	300	60	0	0	250	149	62	0	0	0	0
105	686	55	74	1773	60	66	2551	210	62	3825	256	39	4825	300	60	0	0	250	148	63	0	0	0	0
110	691	60	73	1768	61	65	2562	198	61	3750	255	36	4900	300	60	0	0	250	148	63	0	0	0	0
115	695	65	71	1761	62	71	2575	184	61	3750	253	39	4900	300	60	0	0	250	147	63	0	0	0	0

Table 8, continued /s/ ("s") endpoint, series varying in formant transitions

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
120	701	70	68	1752	59	76	2573	188	61	3750	218	42	4900	300	60	0	0	250	147	63	0	0	0	0
125	708	85	67	1747	64	72	2564	219	60	3809	229	42	4900	300	60	0	0	250	146	63	0	0	0	0
130	718	100	66	1741	69	67	2555	250	59	3867	240	42	4900	300	60	0	0	250	146	63	0	0	0	0
135	727	98	65	1739	70	69	2556	245	59	3883	240	41	4900	300	60	0	0	250	145	63	0	0	0	0
140	736	95	64	1734	71	71	2558	240	59	3900	240	40	4900	300	60	0	0	250	145	63	0	0	0	0
145	747	96	65	1729	70	70	2553	235	60	3866	235	42	4900	300	60	0	0	250	144	63	0	0	0	0
150	758	97	66	1729	68	68	2541	230	60	3885	230	43	4900	300	60	0	0	250	144	63	0	0	0	0
155	769	96	68	1734	82	65	2538	198	59	3903	224	40	4900	300	60	0	0	250	143	63	0	0	0	0
160	777	91	69	1741	96	61	2538	166	57	3909	217	37	4900	300	60	0	0	250	143	63	0	0	0	0
165	779	85	70	1742	96	65	2548	154	59	3914	245	40	4900	300	60	0	0	250	142	62	0	0	0	0
170	778	82	70	1737	95	69	2549	143	60	3932	243	42	4900	300	60	0	0	250	141	62	0	0	0	0
175	776	83	70	1736	98	67	2555	129	57	3957	224	38	4900	300	60	0	0	250	140	62	0	0	0	0
180	774	82	70	1731	100	65	2566	128	54	3974	217	33	4900	300	60	0	0	250	139	61	0	0	0	0
185	773	86	72	1723	96	69	2580	128	56	3987	210	31	4900	300	60	0	0	250	138	61	0	0	0	0
190	772	90	74	1710	92	72	2591	116	58	3986	203	29	4900	300	60	0	0	250	137	61	0	0	0	0
195	769	86	75	1711	89	73	2589	108	59	3963	192	35	4900	300	60	0	0	250	136	60	0	0	0	0
200	761	88	76	1717	87	73	2579	100	59	3940	180	40	4900	300	60	0	0	250	135	60	0	0	0	0
205	750	88	75	1722	81	71	2567	110	59	3978	202	35	4900	300	60	0	0	250	134	58	0	0	0	0
210	737	93	74	1715	76	69	2566	119	59	4016	224	29	4883	300	60	0	0	250	133	56	0	0	0	0
215	725	92	74	1711	77	69	2573	116	58	3993	177	40	4917	300	60	0	0	250	132	54	0	0	0	0
220	712	91	73	1706	78	69	2583	118	56	3952	130	50	4950	300	60	0	0	250	131	52	0	0	0	0
225	704	90	67	1705	83	57	2581	142	50	3922	215	32	4875	300	60	0	0	250	130	50	0	0	0	0
230	695	90	60	1713	88	45	2569	165	44	3870	300	13	4800	300	60	0	0	250	129	48	0	0	0	0
235	691	80	63	1719	104	44	2559	150	48	3818	300	28	4800	300	60	0	0	250	128	51	0	0	0	0
240	679	70	64	1725	120	43	2515	135	52	3769	300	42	4800	300	60	0	0	250	128	53	0	0	0	0
245	677	81	62	1722	98	58	2533	143	45	3736	248	43	4800	300	60	0	0	250	127	47	0	0	0	0
250	672	91	59	1723	76	73	2522	152	37	3710	196	44	4800	300	60	0	0	250	126	40	0	0	0	0

Table 9 /ʃ/ ("sh") endpoint, series varying in formant transitions

Global Parameters:

F Glt Res	B Glt Res	F Glt Zero	B Glt Zero	B Glt Res2
0	100	1500	6000	200
F6	B6	F Nsl Pol	B Nsl Pol	B Nsl Zero
5000	1000	250	100	100
Gain	Auto Amp	No.Cas For	C/P SW	Cor SW
32	-1	5	1	0

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
0	473	45	60	2044	115	54	2712	195	49	3682	160	43	4631	300	37	0	0	250	1	0	0	0	0	65
5	468	45	63	2041	105	55	2716	194	49	3682	145	42	4620	297	37	0	0	250	152	63	0	0	0	0
10	464	45	65	2028	95	57	2724	192	52	3681	130	40	4600	288	37	0	0	250	152	65	0	0	0	0
15	465	39	69	2003	93	61	2729	190	54	3682	115	42	4580	274	39	0	0	250	153	66	0	0	0	0
20	472	42	65	1973	91	64	2730	198	56	3686	100	43	4572	258	41	0	0	250	142	69	0	0	0	0
25	483	45	65	1948	80	65	2730	200	55	3691	93	43	4588	246	42	0	0	250	143	67	0	0	0	0
30	505	48	65	1917	72	64	2735	203	54	3693	103	43	4633	238	42	0	0	250	144	73	0	0	0	0
35	534	54	64	1881	70	62	2736	198	53	3689	125	43	4697	232	39	0	0	250	145	73	0	0	0	0
40	563	56	65	1846	73	65	2736	193	58	3672	156	47	4755	219	45	0	0	250	147	70	0	0	0	0
45	586	60	68	1825	81	71	2730	188	69	3648	186	55	4800	200	60	0	0	250	147	70	0	0	0	0
50	591	57	71	1829	86	76	2731	170	66	3610	177	55	4783	200	60	0	0	250	148	66	0	0	0	0
55	599	55	71	1820	85	75	2725	151	64	3605	185	50	4750	200	60	0	0	250	148	65	0	0	0	0
60	606	57	70	1811	78	74	2710	140	62	3599	226	44	4716	200	60	0	0	250	149	65	0	0	0	0
65	612	60	70	1802	69	71	2685	153	64	3625	263	44	4695	225	60	0	0	250	149	66	0	0	0	0
70	620	64	68	1799	61	67	2658	174	65	3650	288	44	4673	250	60	0	0	250	149	68	0	0	0	0
75	629	55	74	1798	57	70	2643	186	65	3725	281	42	4677	275	60	0	0	250	149	65	0	0	0	0
80	640	45	80	1796	55	72	2623	210	65	3800	275	42	4681	300	60	0	0	250	150	65	0	0	0	0
85	655	41	80	1792	60	70	2605	211	66	3767	268	42	4685	300	60	0	0	250	149	61	0	0	0	0
90	670	37	79	1787	65	67	2578	212	66	3733	261	42	4689	300	60	0	0	250	149	62	0	0	0	0
95	675	44	77	1782	61	67	2562	212	64	3817	259	42	4719	300	60	0	0	250	149	63	0	0	0	0
100	681	50	75	1778	57	66	2551	213	62	3900	258	42	4750	300	60	0	0	250	149	64	0	0	0	0
105	686	55	74	1773	60	66	2551	210	62	3825	256	39	4825	300	60	0	0	250	148	65	0	0	0	0
110	691	60	73	1768	61	65	2562	198	61	3750	255	36	4900	300	60	0	0	250	148	65	0	0	0	0
115	695	65	71	1761	62	71	2575	184	61	3750	253	39	4900	300	60	0	0	250	147	67	0	0	0	0

Table 9, continued /ʃ/ ("sh") endpoint, series varying in formant transitions

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
120	701	70	68	1752	59	76	2573	188	61	3750	218	42	4900	300	60	0	0	250	147	67	0	0	0	0
125	708	85	67	1747	64	72	2564	219	60	3809	229	42	4900	300	60	0	0	250	146	69	0	0	0	0
130	718	100	66	1741	69	67	2555	250	59	3867	240	42	4900	300	60	0	0	250	146	69	0	0	0	0
135	727	98	65	1739	70	69	2556	245	59	3883	240	41	4900	300	60	0	0	250	145	69	0	0	0	0
140	736	95	64	1734	71	71	2558	240	59	3900	240	40	4900	300	60	0	0	250	145	69	0	0	0	0
145	747	96	65	1729	70	70	2553	235	60	3866	235	42	4900	300	60	0	0	250	144	69	0	0	0	0
150	758	97	66	1729	68	68	2541	230	60	3885	230	43	4900	300	60	0	0	250	144	69	0	0	0	0
155	769	96	68	1734	82	65	2538	198	59	3903	224	40	4900	300	60	0	0	250	143	67	0	0	0	0
160	777	91	69	1741	96	61	2538	166	57	3909	217	37	4900	300	60	0	0	250	143	67	0	0	0	0
165	779	85	70	1742	96	65	2548	154	59	3914	245	40	4900	300	60	0	0	250	142	66	0	0	0	0
170	778	82	70	1737	95	69	2549	143	60	3932	243	42	4900	300	60	0	0	250	141	66	0	0	0	0
175	776	83	70	1736	98	67	2555	129	57	3957	224	38	4900	300	60	0	0	250	140	66	0	0	0	0
180	774	82	70	1731	100	65	2566	128	54	3974	217	33	4900	300	60	0	0	250	139	65	0	0	0	0
185	773	86	72	1723	96	69	2580	128	56	3987	210	31	4900	300	60	0	0	250	138	65	0	0	0	0
190	772	90	74	1710	92	72	2591	116	58	3986	203	29	4900	300	60	0	0	250	137	65	0	0	0	0
195	769	86	75	1711	89	73	2589	108	59	3963	192	35	4900	300	60	0	0	250	136	62	0	0	0	0
200	761	88	76	1717	87	73	2579	100	59	3940	180	40	4900	300	60	0	0	250	135	59	0	0	0	0
205	750	88	75	1722	81	71	2567	110	59	3978	202	35	4900	300	60	0	0	250	134	61	0	0	0	0
210	737	93	74	1715	76	69	2566	119	59	4016	224	29	4883	300	60	0	0	250	133	61	0	0	0	0
215	725	92	74	1711	77	69	2573	116	58	3993	177	40	4917	300	60	0	0	250	132	60	0	0	0	0
220	712	91	73	1706	78	69	2583	118	56	3952	130	50	4950	300	60	0	0	250	131	59	0	0	0	0
225	704	90	67	1705	83	57	2581	142	50	3922	215	32	4875	300	60	0	0	250	130	58	0	0	0	0
230	695	90	60	1713	88	45	2569	165	44	3870	300	13	4800	300	60	0	0	250	129	58	0	0	0	0
235	691	80	63	1719	104	44	2559	150	48	3818	300	28	4800	300	60	0	0	250	128	57	0	0	0	0
240	679	70	64	1725	120	43	2515	135	52	3769	300	42	4800	300	60	0	0	250	128	56	0	0	0	0
245	677	81	62	1722	98	58	2533	143	45	3736	248	43	4800	300	60	0	0	250	127	56	0	0	0	0
250	672	91	59	1723	76	73	2522	152	37	3710	196	44	4800	300	60	0	0	250	126	55	0	0	0	0

Table 10 /*ʃ/ (beyond "sh") endpoint, series varying in formant transitions

Global Parameters:

F Glt Res	B Glt Res	F Glt Zero	B Glt Zero	B Glt Res2
0	100	1500	6000	200
F6	B6	F Nsl Pol	B Nsl Pol	B Nsl Zero
5000	1000	250	100	100
Gain	Auto Amp	No.Cas For	C/P SW	Cor SW
32	-1	5	1	0

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
0	432	50	41	2404	145	52	2672	265	39	3654	120	37	4401	300	34	0	0	250	1	0	0	0	0	65
5	428	45	46	2391	124	49	2676	265	39	3645	95	35	4400	295	62	0	0	250	152	63	0	0	0	0
10	404	40	48	2368	105	45	2696	265	42	3621	64	27	4410	278	40	0	0	250	152	65	0	0	0	0
15	399	24	57	2322	103	47	2709	257	45	3621	30	32	4567	280	19	0	0	250	153	66	0	0	0	0
20	383	26	48	2267	101	49	2714	278	50	3626	1	33	4545	280	25	0	0	250	142	69	0	0	0	0
25	383	31	50	2214	76	53	2720	290	45	3638	38	36	4528	249	25	0	0	250	143	67	0	0	0	0
30	405	38	56	2147	56	53	2728	300	43	3687	75	40	4567	228	25	0	0	250	144	73	0	0	0	0
35	474	50	57	2011	60	53	2739	248	41	3685	112	35	4648	238	19	0	0	250	145	76	0	0	0	0
40	537	53	65	1871	70	63	2726	190	48	3659	149	41	4714	239	31	0	0	250	147	70	0	0	0	0
45	586	60	68	1825	81	71	2730	188	69	3648	186	55	4800	200	60	0	0	250	147	70	0	0	0	0
50	591	57	71	1829	86	76	2731	170	66	3610	177	55	4783	200	60	0	0	250	148	67	0	0	0	0
55	599	55	71	1820	85	75	2725	151	64	3605	185	50	4750	200	60	0	0	250	148	66	0	0	0	0
60	606	57	70	1811	78	74	2710	140	62	3599	226	44	4716	200	60	0	0	250	149	66	0	0	0	0
65	612	60	70	1802	69	71	2685	153	64	3625	263	44	4695	225	60	0	0	250	149	67	0	0	0	0
70	620	64	68	1799	61	67	2658	174	65	3650	288	44	4673	250	60	0	0	250	149	68	0	0	0	0
75	629	55	74	1798	57	70	2643	186	65	3725	281	42	4677	275	60	0	0	250	149	65	0	0	0	0
80	640	45	80	1796	55	72	2623	210	65	3800	275	42	4681	300	60	0	0	250	150	65	0	0	0	0
85	655	41	80	1792	60	70	2605	211	66	3767	268	42	4685	300	60	0	0	250	149	61	0	0	0	0
90	670	37	79	1787	65	67	2578	212	66	3733	261	42	4689	300	60	0	0	250	149	62	0	0	0	0
95	675	44	77	1782	61	67	2562	212	64	3817	259	42	4719	300	60	0	0	250	149	63	0	0	0	0
100	681	50	75	1778	57	66	2551	213	62	3900	258	42	4750	300	60	0	0	250	149	64	0	0	0	0
105	686	55	74	1773	60	66	2551	210	62	3825	256	39	4825	300	60	0	0	250	148	65	0	0	0	0
110	691	60	73	1768	61	65	2562	198	61	3750	255	36	4900	300	60	0	0	250	148	65	0	0	0	0
115	695	65	71	1761	62	71	2575	184	61	3750	253	39	4900	300	60	0	0	250	147	66	0	0	0	0

Table 10, continued /*ʃ/ (beyond "sh") endpoint, series varying in formant transitions

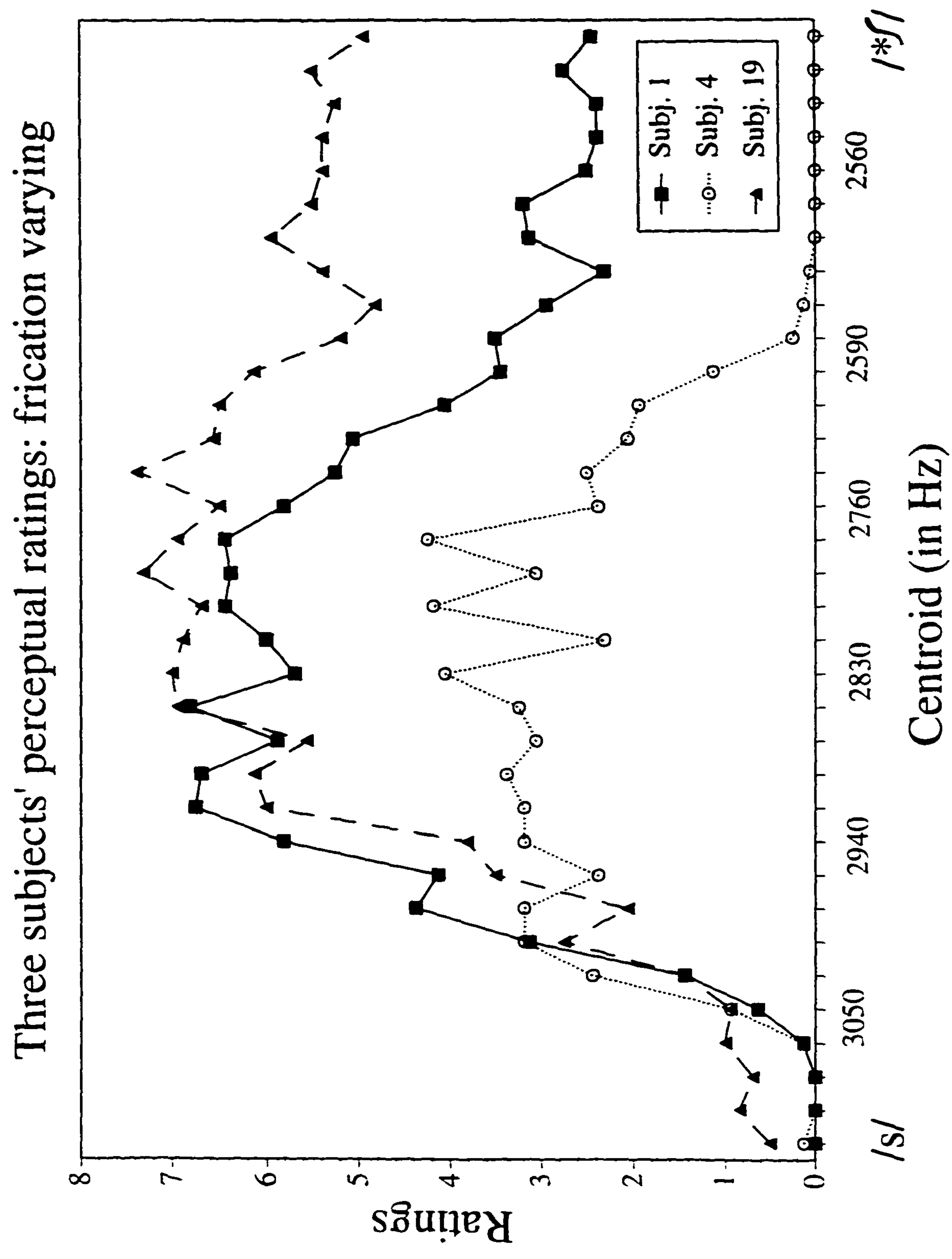
msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
120	701	70	68	1752	59	76	2573	188	61	3750	218	42	4900	300	60	0	0	250	147	66	0	0	0	0
125	708	85	67	1747	64	72	2564	219	60	3809	229	42	4900	300	60	0	0	250	146	68	0	0	0	0
130	718	100	66	1741	69	67	2555	250	59	3867	240	42	4900	300	60	0	0	250	146	69	0	0	0	0
135	727	98	65	1739	70	69	2556	245	59	3883	240	41	4900	300	60	0	0	250	145	69	0	0	0	0
140	736	95	64	1734	71	71	2558	240	59	3900	240	40	4900	300	60	0	0	250	145	69	0	0	0	0
145	747	96	65	1729	70	70	2553	235	60	3866	235	42	4900	300	60	0	0	250	144	68	0	0	0	0
150	758	97	66	1729	68	68	2541	230	60	3885	230	43	4900	300	60	0	0	250	144	68	0	0	0	0
155	769	96	68	1734	82	65	2538	198	59	3903	224	40	4900	300	60	0	0	250	143	67	0	0	0	0
160	777	91	69	1741	96	61	2538	166	57	3909	217	37	4900	300	60	0	0	250	143	67	0	0	0	0
165	779	85	70	1742	96	65	2548	154	59	3914	245	40	4900	300	60	0	0	250	142	66	0	0	0	0
170	778	82	70	1737	95	69	2549	143	60	3932	243	42	4900	300	60	0	0	250	141	66	0	0	0	0
175	776	83	70	1736	98	67	2555	129	57	3957	224	38	4900	300	60	0	0	250	140	66	0	0	0	0
180	774	82	70	1731	100	65	2566	128	54	3974	217	33	4900	300	60	0	0	250	139	66	0	0	0	0
185	773	86	72	1723	96	69	2580	128	56	3987	210	31	4900	300	60	0	0	250	138	65	0	0	0	0
190	772	90	74	1710	92	72	2591	116	58	3986	203	29	4900	300	60	0	0	250	137	65	0	0	0	0
195	769	a6	75	1711	89	73	2589	108	59	3963	192	35	4900	300	60	0	0	250	136	62	0	0	0	0
200	761	88	76	1717	87	73	2579	100	59	3940	180	40	4900	300	60	0	0	250	135	60	0	0	0	0
205	750	88	75	1722	81	71	2567	110	59	3978	202	35	4900	300	60	0	0	250	134	61	0	0	0	0
210	737	93	74	1715	76	69	2566	119	59	4016	224	29	4a83	300	60	0	0	250	133	61	0	0	0	0
215	725	92	74	1711	77	69	2573	116	58	3993	177	40	4917	300	60	0	0	250	132	60	0	0	0	0
220	712	91	73	1706	78	69	2583	118	56	3952	130	50	4950	300	60	0	0	250	131	59	0	0	0	0
225	704	90	67	1705	83	57	2581	142	50	3922	215	32	4875	300	60	0	0	250	130	58	0	0	0	0
230	695	90	60	1713	88	45	2569	165	44	3870	300	13	4800	300	60	0	0	250	129	58	0	0	0	0
235	691	80	63	1719	104	44	2559	150	48	3818	300	28	4800	300	60	0	0	250	128	57	0	0	0	0
240	679	70	64	1725	120	43	2515	135	52	3769	300	42	4800	300	60	0	0	250	128	56	0	0	0	0
245	677	81	62	1722	98	58	2533	143	45	3736	248	43	4800	300	60	0	0	250	127	56	0	0	0	0
250	672	91	59	1723	76	73	2522	152	37	3710	196	44	4800	300	60	0	0	250	126	55	0	0	0	0

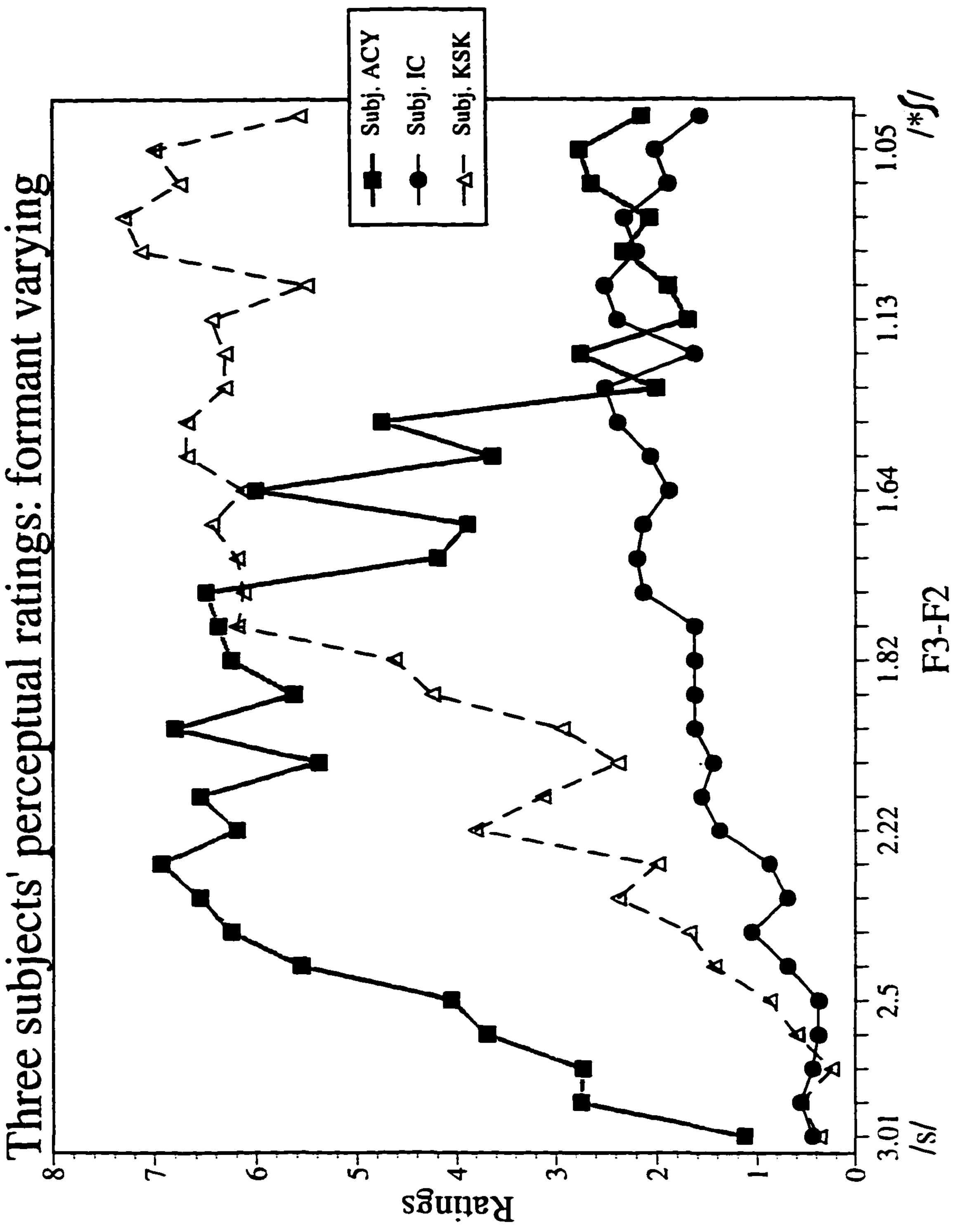
total of 56 trials in this block (4 tokens x 2 consonants x 7 vowel environments). The program was then run a second time, so that each subject recorded eight tokens of each CV syllable.

For the perception task, half of the subjects listened to the items varying in frication centroid first (that is, during session 1 and the first half of session 2), and half listened to the items varying in formant values first. Session 1 included the production component, and 10 blocks of trials in the perceptual experiment (during which listeners heard either the items varying in frication centroid or those varying in formant frequency values). Session 2 consisted of 6 blocks of each of the two series (or a total of 12 blocks), and session 3 consisted of the remaining 10 blocks of perceptual trials. As in Experiment 1, subjects were asked to rate the initial phoneme for its goodness as an example of the category /ʃ/. Subjects responded using the numbers zero through nine on a numeric keypad, followed by the "return" or "enter" key. Subjects were told to use the "0" label whenever the item did not sound like an "sh" at all, to use the "1" whenever it was unclear whether it was an "sh" or not, and to use the range "2" through "9" for items which were definitely members of the category "sh", but differed in how good of an example they were. Subjects were given a reference sheet which contained this scale, in case they wished to refer back to it. While subjects' response times were not recorded, they were informed that the next trial would begin as soon as they responded to the current trial.

Results and Discussion

Results were measured as in the first experiment. For the perception task, the single item in each continuum (F3 - F2 values varying and frication centroid varying) with the highest rating was considered the listener's prototype for that dimension. Figure 5 shows the rating functions for the frication-varying series for three subjects who participated in this experiment. Figure 6 shows the rating functions for the formant-varying series for three participants. As in Experiment 1, the subjects' ratings generally increased until they reached a peak, and then began decreasing, leaving a single item as a prototype. As in Experiment 1, although some individuals had 2, or possibly 3, items which received very similar ratings, the single item with the highest rating was selected as their prototype. Given the slight acoustic differences between adjacent members of the series, this is unlikely to result in large amounts of noise. Furthermore, the subjects' prototypes varied over a moderately large range (for centroids, 2739 - 2935 Hz; for formant differences, 0.83 - 2.24 Bark), such that this small amount of potential noise in prototype selection is unlikely to change the overall results.


Three subjects' perceptual ratings: frication varying
Ratings
0
1
2
3
4
5
6
7
8
/s/
3050
2940
2830
2760
2590
2560
/*ʃ/
Centroid (in Hz)
Subj. 1
Subj. 4
Subj. 19


Three subjects' perceptual ratings: formant varying
Ratings
0
1
2
3
4
5
6
7
8
F3-F2
3.01
2.5
2.22
1.82
1.64
1.13
1.05
/s/
/*ʃ/
Subj. ACY
Subj. IC
Subj. KSK

For the production experiment, F3, F2 and the frication centroid were measured for each token. Frication centroids are really an amplitude-weighted mean frequency value of the energy present in the fricative spectrum. That is, a cross-section of the fricative at one moment in time is taken, and from this the amount of energy present at each frequency is determined. This is treated as a distribution, and from this distribution it is possible to find a mean or average frequency value. Frequency centroids were computed with 15-ms segments (or frames) of the waveform and repeated every 5 ms over the stimulus. These values were then averaged over the first 20 frames of each stimulus. Thus, the mean calculation was based on information over the first 100 ms of the frication.[16] This duration was chosen because Tomiak (1991) suggested it as a valid estimate based on results from a masking study. Although other researchers have made different choices in this regard, these differences in methodology are unlikely to result in substantial differences. For instance, Behrens and Blumstein (1988) examined three separate 15 ms windows, one at the onset of frication, one at the end of frication, and one in the middle of the frication, and found that their peak measures were relatively constant across time.

---

[16] There were four productions where the noise portion was shorter than 100 ms. In these cases, calculations were averaged across 15 frames (or 75 ms.)

F3 and F2 were also measured using a 15 ms temporal window. The window was centered on the first vocal pulse, and the measurements from this and the following two 5-ms time frames were averaged to get a more reliable estimate of the formant frequencies. These were then transformed into Bark scale equivalents (Zwicker & Terhardt, 1980), and F2 was subtracted from F3. Values for frequency centroid and for F3 - F2 were averaged across the eight tokens of each intended syllable.

As there were fewer items of interest in this experiment than in Experiment 1, single correlations were used rather than multiple regressions. To control for an increased number of statistical tests, alpha levels of .01 were used instead of .05.

For the frication centroid-varying series, there were no significant correlations. For the production measures on the syllable /ʃæ/, the correlation with the perceptual prototype was -.26 (z=-1.10, $p \geq .27$). Including all /ʃ/ productions, the correlation was approximately equivalent (r=-.25, z=-1.04, $p \geq .29$). For the /s/ productions, the correlation was even lower (r=-.02, z=-0.08, $p \geq .94$). Thus, there does not seem to be any correlation between the centroids of frication in subjects' productions and in their perceptual prototypes.

For the series varying in formant values, there were likewise no significant correlations. For the production measures on the syllable /ʃæ/, the correlation with perception was -.28 (z=-1.20, $p \geq .23$). For all /ʃ/ productions, the correlation was even lower (r= -.17; z=-0.70, $p \geq .48$), while for /s/ productions it was statistically marginal, but in the wrong direction ($r = -.49$, z=-2.18, $p < .03$). Thus, there does not appear to be strong evidence for a correlation in the formant values of tokens subjects produced and the values for subjects' perceptual prototypes.

Given our results from the first Experiment, this lack of an effect is somewhat surprising. There are a number of possible reasons for this. One potential problem with the production results is that the listeners may have been mimicking the talker they heard, even though they were explicitly instructed to produce the items normally. Goldinger (1997) has found that listeners in a shadowing task tend to mimic the speakers they hear. It is not clear why this group of subjects would have done so when the group of subjects in Experiment 1 did not. However, it is possible that the specific design of this experiment encouraged listeners to pay more attention to between-token differences than they did in the previous experiment. In Experiment 1, they only heard each CV once in each block (with the exception of the target CV, which they heard three times). Here, they heard each CV four times. Furthermore, there were only two possible consonants in this experiment, rather than the six in the first experiment. Since the participants were hearing each syllable several times, and hearing each consonant even more times, they may have begun to pay attention to particular aspects of the way in which the syllables were produced, and begun mimicking these idiosyncrasies. In order to investigate this potential confound, the variability of the original talker's productions of CVs with low-vowels were examined. Low vowels were chosen because it was predicted that there would be more room for consonant variability in these cases than there would be for high vowels (which have very extreme formant values; these extreme values may place limitations on the amount of variability that could be found in the consonant, as the talker would need to be moving towards the formant values for the vowel at an earlier point in time). Upon investigation, it was discovered that the talker's productions of /ʃo/ contained the most variability, with centroids ranging from 4987 Hz to 5266 Hz. Subjects' productions were then examined for this same syllable. If participants were mimicking the talker, then they should have produced higher centroids for /ʃo/ after hearing the token with the 5266 Hz centroid, and produced lower centroids after hearing the token with a centroid value of 4987 Hz. That is, they should have shown the same pattern of centroid production as the talker, producing higher centroids when her token contained higher centroids, and producing lower centroids when her tokens contained lower centroids. A paired t-test was performed on the centroid values of participants' productions following the tokens in which the talker (RSN) had produced the highest and lowest centroid values. No significant difference was found in participants' productions following these two example tokens (t =0.84, p >.40).

To investigate whether there may have been a trend towards mimicking the talker that was not large enough to produce significant differences, the centroid values across talkers for all four /ʃo/ tokens were examined. If participants' productions were influenced by the values of the token they heard, their productions overall should have resulted in the same rank-order as the original talker's productions. RSN's centroid values were 5266, 4987, 5240, and 5238. Thus, the rank ordering for her tokens (from lowest to highest) would be 2, 4, 3, and 1 (that is, her second token had the lowest centroid, than her 4th and 3rd, and her first token had the highest centroid. As the intermediate two, tokens 3 and 4, were approximately equal, their ordering relative to one another might be expected to change. However, they should still be ordered intermediate to the first and second productions). The participants' productions did not follow the same pattern. Their average values were 5186, 5083, 5193, and 5200, and thus their ordering would be 2, 1, 3, and 4. Thus, the ordering of subjects' productions did not follow the ordering of the talker's productions. Combined with the nonsignificant difference from the t-test, it does not appear likely that listeners were mimicking the talker they heard to any great degree.

Another possible explanation for the null result is that the notion that the degree of production-perception correlation is related to the extent the measure is appropriate (or the extent to which it is correlated with the dimensions actually used by the subjects) may not be correct. Certainly this would have been the conclusion had the correlations for the secondary (formant-based) cue been larger than the correlations for the primary (frication-based) cue. However, given that both cues led to null results, this argument loses some of its force. Still, this possibility can not be ruled out.

A third potential explanation is that overall mean may not be the most accurate cue to frication. Although a great deal of research suggests that the energy during frication is the primary cue to the /s/-/ʃ/ distinction, the centroid, or mean value, may not be the most appropriate way of measuring this. Several researchers (Jassem, 1965; Behrens & Blumstein, 1988) have examined peaks in the frication spectrum, rather than overall centroids. While these two measures would be identical if frication noise formed a normal distribution of energy, this is not necessarily the case. A peak in frication energy is more akin to the statistical "mode", rather than the "mean", and the mean (or centroid) will be influenced to a much greater extent by low amplitude, high frequency energy (akin to statistical "outliers"). Results from Behrens and Blumstein (1988) and Jassem (1965) suggest that peak values for /ʃ/ range from 2.5-3.5 kHz, whereas peak values for /s/ range between 3.5 and 5 kHz. With 10 kHz stimuli, this results in a greater potential for extremely high frequency energy than there is for extremely low frequency energy (as there can be no energy below 0 kHz). That is, there is a more limited range of potential outliers at the lower frequencies than at higher frequencies. This is likely to result in a somewhat skewed distribution of frication energy, and thus for a sizable difference between centroid (mean) and peak (mode) values. If subjects are relying more heavily on peak information than on centroid information, this could easily result in the null results found here.

Yet another possibility is that listeners do calculate mean values, but do so within different frequency bands, rather than computing an overall centroid. Although such a notion has not been formally proposed, it would be consistent with much of the previous literature. In the present experiment, the perceptual data were based on stimuli that only contained frequency information as high as 5 kHz, whereas individual's production values included energy as high as 10 kHz. It is possible that some listeners whose mean production values were quite high may have had their means heavily influenced by information above 5 kHz. In fact, within the range of 0-5 kHz, their mean values might have actually been lower, on average, than were the productions of individuals whose overall means were less high. That is, some individuals might prefer to produce /s/ sounds with less energy in the 4-5 kHz range, but compensate for this by producing energy above 5 kHz. Given a perceptual task in which the sounds they heard only had energy as high as 5 kHz, their prototype would appear relatively low in overall mean.

Although this explanation is very *post hoc*, an examination of a few of the subjects' productions down-sampled to 5 kHz produced some very interesting results. Specifically, subjects' productions of /s/ and /ʃ/ did not appear to differ on their overall mean frequency within this more limited frequency range, even though they were perceptually distinct. That is, mean frequency did not seem to work as a cue in down-sampled speech. Yet, we rarely have difficulty understanding individuals on the telephone, even though telephones do not carry acoustic information above 5 kHz. Lexical context likely plays a large role in this situation, but context cannot assist in the perception of peoples' names. Names are often difficult to understand on the telephone, but rarely impossible. Since people can still distinguish /s/ and /ʃ/ productions without higher-frequency information, even if the mean values no longer differ, it is suggestive that mean frequency of frication may not be the cue listeners are actually using.

One odd finding was that all six correlations, although nonsignificant, were consistently negative. While it is possible that this is meaningless, it is also possible that this indicates a trend of some sort, albeit in an unexpected direction. It is unclear why these correlations would be negative, although some recent work with vowels has shown a similar pattern of findings (Frieda, 1997). The negative correlation suggests that individuals who produced relatively high centroids of frication preferred hearing lower centroids, while individuals with relatively low centroids preferred higher ones. One possibility is that this might be driven by individuals with more extreme values. If an individual X realizes that his own "s" productions are aberrantly high in frequency, he might take this into account when attempting to rate another talker's "s" productions. This would cause him to rate lower the talker's "s" items that are closer to his own productions, and to rate more "normal" productions more highly. If subjects with relatively low centroids did likewise, this could result in a crossing effect, with high-frequency individuals having lower-frequency prototypes than low-frequency individuals. Although plausible, this explanation is entirely *post hoc*, and must be viewed with some skepticism until future research can examine it in more detail.

One last issue concerns the hyperarticulation effect discussed in Experiment 1. Individuals who participated in the first experiment had perceptual prototypes that were more extreme than their own productions. Unfortunately, it is not possible to determine whether a similar difference occurs with frication centroids. This is because the synthetic stimuli used in the perception experiment only contained energy at frequencies below 5 kHz, whereas the individuals' productions included energy at frequencies up to approximately 10 kHz. Since frication centroids are sensitive to this high-frequency information, the production tokens almost by necessity have higher centroids than the perceptual tokens. This makes it impossible to determine whether there was any difference between perception and production caused by a preference for hyperarticulated tokens.

It is possible to examine this with the formant-varying series, however. Listeners reliably preferred items that had smaller F3 - F2 differences than they produced in their own tokens ($t = 5.323$, $p < .0001$). That is, listeners preferred for the formants to be closer together. However, as formant differences are generally considered to be a secondary cue, this result might be an artifact of the testing situation. In order to make a series that varied in category goodness, the frication portion of the formant-varying series was somewhat ambiguous. (This was done to avoid the frication being such a salient cue as to overwhelm the potentially lesser cue of formant structure.) This may have forced listeners to pay close attention to the secondary cue, and perhaps to depend more on this cue in the perceptual task than they did in the production task. If participants depended primarily on frication in their production, but were forced (due to an ambiguous frication) to depend more heavily on formant structure in perception, this same effect would have occurred. If the subjects were marking the /s/-/ʃ/ distinction in production primarily by the frication, there would be no need for them to vary the formant structures in a distinctive manner. In the perceptual task, the frication cue was nondistinctive, so listeners had no choice but to rate items on the basis of these formant differences. Inevitably, then, the formant differences would be more distinctive in the results from the perception task than from the production task. Thus, the apparent hyperarticulation effect may in fact be due solely to the specific demands of this experiment. It may not be appropriate to search for effects of hyperarticulation in conditions focusing on non-primary cues.

In conclusion, the present experiment does not provide further evidence for the existence of a link between perception and production. It is at least possible that the failure to this result may have been caused by examining an inappropriate cue. However, further research will be needed to examine this possibility in more depth.

Given the positive results from Experiment 1, however, it appears that it is at least possible to find perception-production correlations in some circumstances. Perhaps this methodology could be used to evaluate different proposed perceptual cues. Often, there are multiple proposals for how a given phonemic distinction should be described. It might be possible to evaluate different metrics by determining the degree to which perception and production measures using these proposed cues are correlated. Experiment 3 describes this in more detail.

CHAPTER 4

Experiment 3: A comparison of different metrics

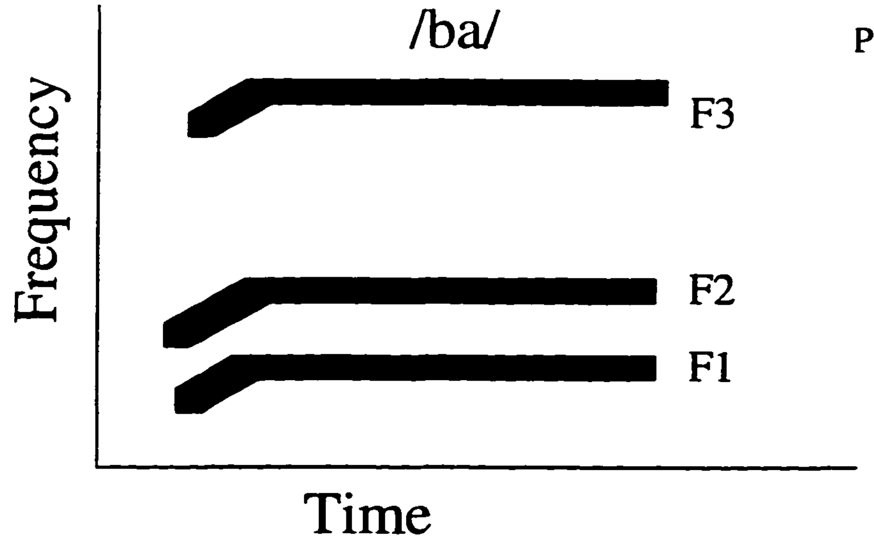
Unlike the /p/-/b/ distinction discussed in Chapter 2, there are some phonemic distinctions (such as place-of-articulation in stops) where many different metrics appear to be equally plausible. One reason for this multitude of proposals is that the acoustic spectrum for these phonemes is rather complex, and the differences between spectra can be described in a number of ways.

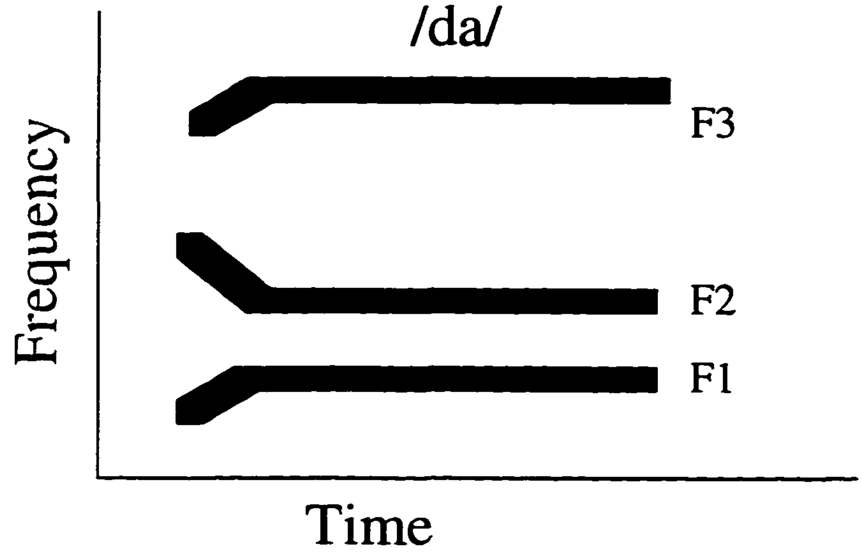
As discussed in Chapter 2, when speakers produce stop consonants, they create an obstruction in the mouth, blocking air flow. Air pressure builds up in the oral cavity and then is released explosively. At some point thereafter, the vocal folds begin to vibrate. The time delay between these two events distinguishes the "voiced" stops (b, d, and g, which have short delays) from the "voiceless" stops (p, t, and k, which have long delays). However, the acoustic cues distinguishing between /b/, /d/, and /g/ (or between /p/, /t/, and /k/) are less obvious.

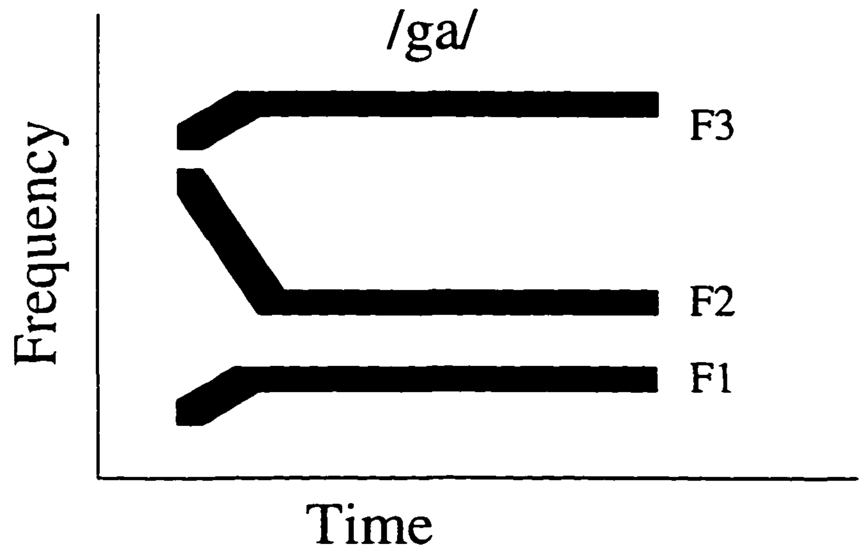
In terms of articulation, the "b" is produced by causing an obstruction at the lips. The "d" is produced by pressing the tongue against the alveolar ridge (the section immediately behind the teeth in the top of the mouth). The "g" is produced further back in the mouth, by pressing the blade of the tongue against the hard palate (the roof of the mouth).

As described in Chapter 3, the location of the tongue, jaw, etc. changes the shape of the vocal tract. This emphasizes different frequencies in the signal. With the stop consonants, the occlusion divides the vocal tract into two portions. As the occlusion is moved further back in the mouth, the portion before the obstruction becomes smaller, and the portion following the constriction becomes larger. These changes cause different frequencies to be emphasized, both in the burst (at the release of air pressure) and once the vocal fold vibration begins. When the obstruction is released, the tongue (or lips) moves rapidly away from the location of the constriction and into whatever position is necessary for the following vowel. This causes a rapid change in the formants (that is, in the frequencies that get emphasized by the vocal tract). This is apparent in Figure 7 which shows a schematic diagram of the formant locations for /b/, /d/, and /g/. The formants move sharply at the onset of the syllables, as the tongue and jaw move away from the occlusion position and into position for the following vowel.

It is well-known that the information in these spectrum correlates with the location of the articulators in the mouth, and thus can be an indication of the sound the speaker intended to produce. What is less clear is the best way to describe (or condense) this information. Since the formants are dependent on the shape of the vocal tract, the exact frequency


/ba/
F3
F2
F1
Frequency
Time


/da/
F3
F2
F1
Frequency
Time

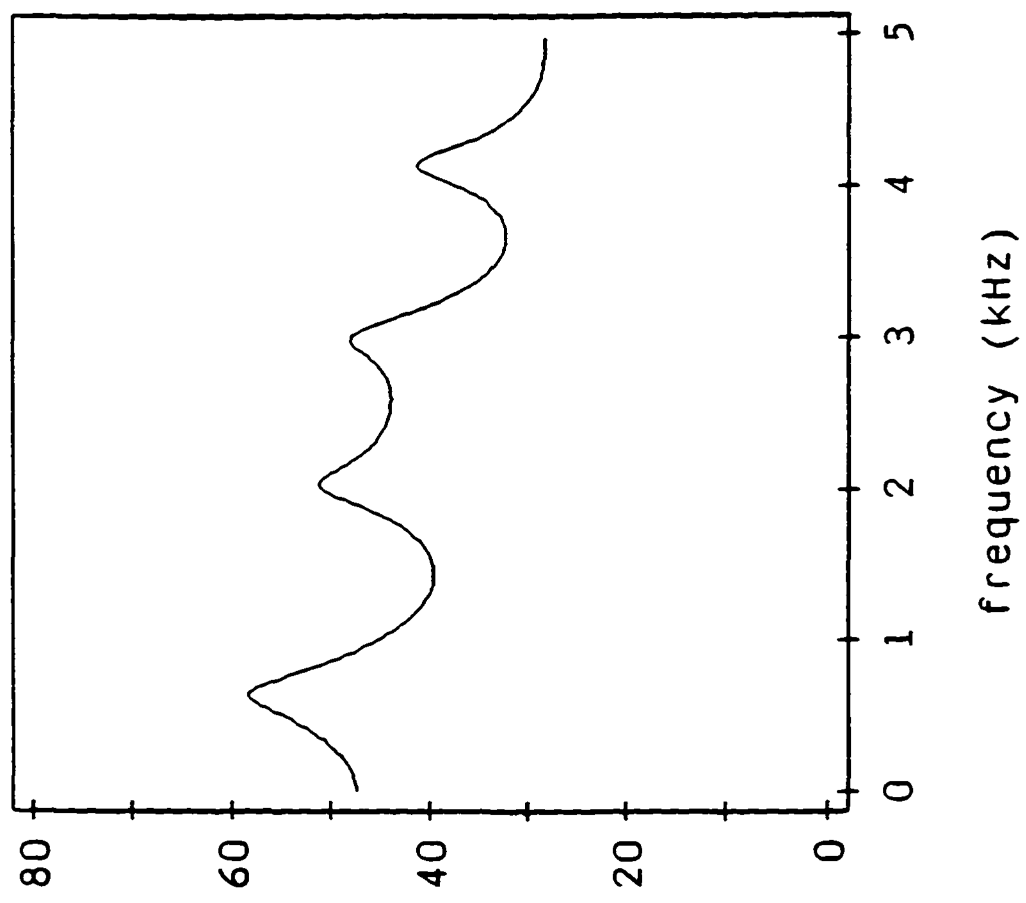

/ga/
F3
F2
F1
Frequency
Time

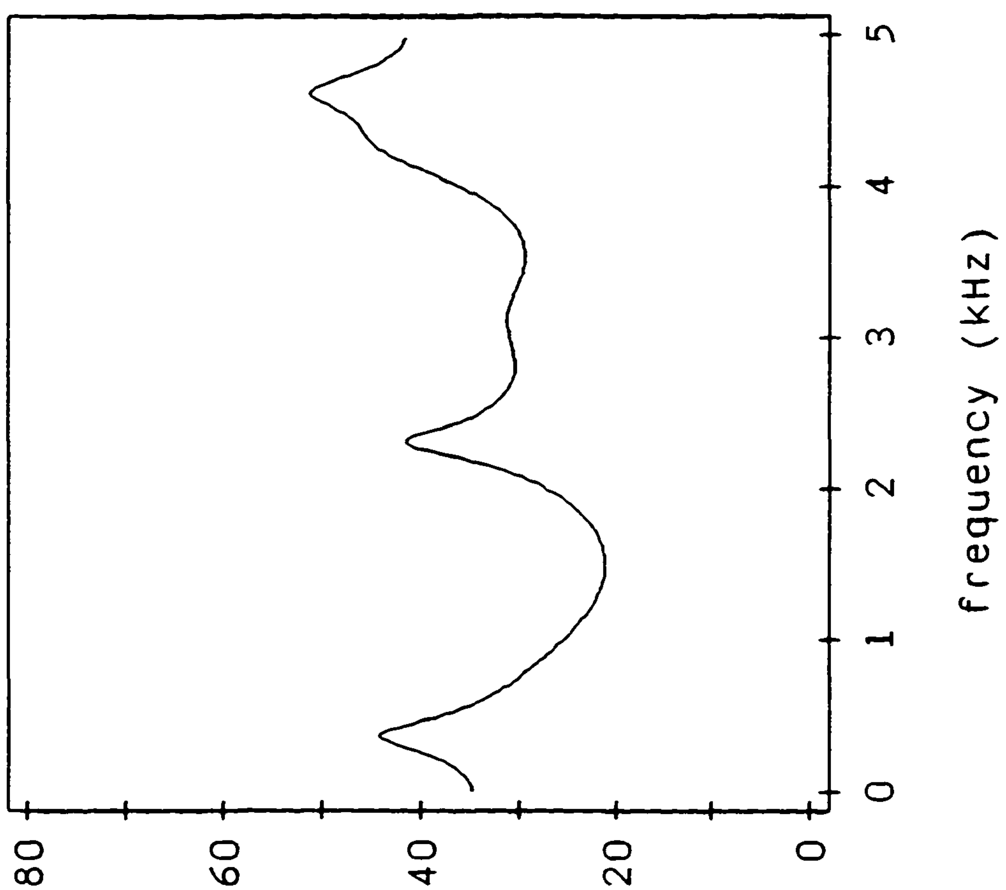
values will be different for different people. Thus, people with larger vocal tracts will have lower formants, and people with smaller vocal tracts will have higher formants. This means the exact values of these formants are not an invariant cue, and researchers have struggled to find ways of describing the spectrum that are less variable with differences between talkers.

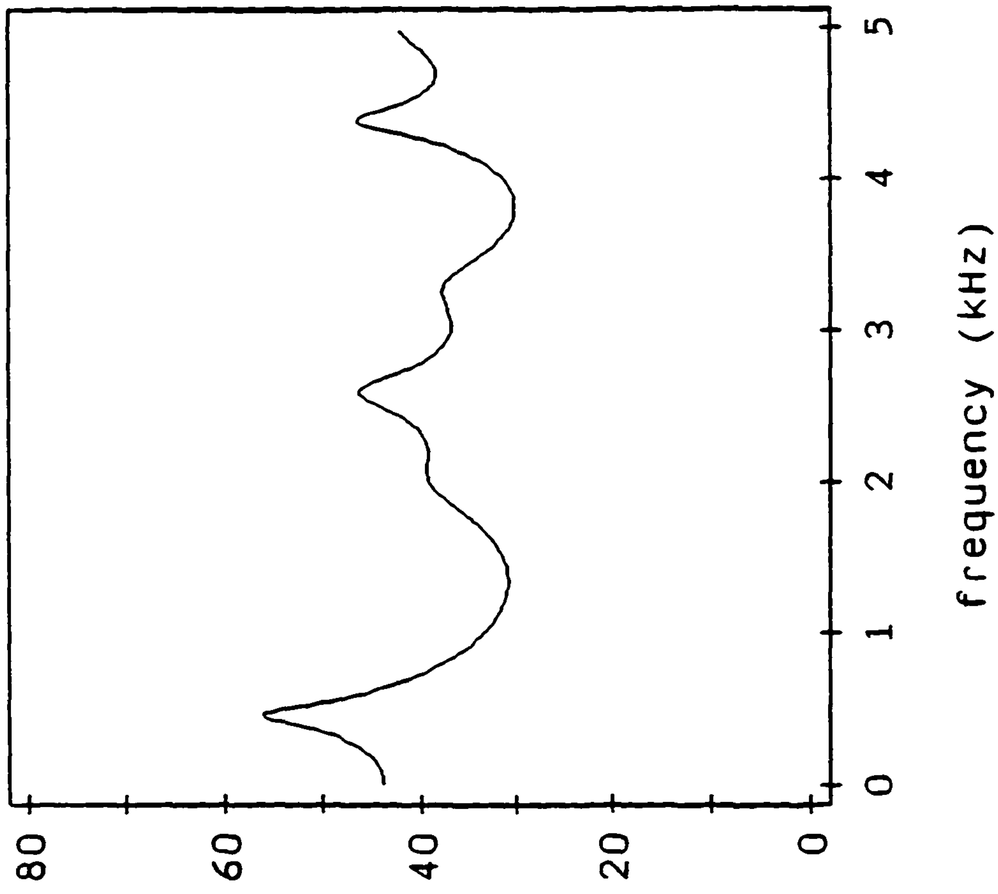
This has led researchers to propose a variety of possible cues, and then to examine whether algorithms based on these proposals classify sounds in the same manner as do human listeners. One example is spectral tilt (the shape of the short-term spectra at onset), which was first described by Stevens and Blumstein (1978; Blumstein & Stevens, 1979). Other proposed metrics are based on spectral moments (the mean, variability, skewness, and kurtosis of the energy distribution; Forrest et al., 1988; Sawusch & Dutton, 1992), peak differences (the distances between bands of energy that are emphasized by the vocal tract; Syrdal & Gopal, 1986), and F2 locus equations (the starting point of the second formant; Sussman, McCaffrey & Matthews, 1991; Sussman, Hoemeke & Ahmed, 1993; Sussman, 1991; Sussman, 1989). It has been difficult to distinguish among these different proposals experimentally. Examining the perception/production correlations for different metrics may provide a new way of doing so.

The metrics

Spectral tilt. The original version of this proposal was that the gross shape of the short-term spectra at onset was invariant for place-of-articulation (that is, the location of the obstruction in the mouth could be determined by the general distribution of energy across the frequency range). Stevens and Blumstein (1978; Blumstein & Stevens, 1979; 1980) suggested that this cue contained information from both the formant transitions and the burst, rather than from either one alone (as did many other suggested invariants). Bilabial stops are characterized by a diffuse, falling spectrum, alveolars by a diffuse rising spectrum, and velars by a compact spectrum. That is, bilabials (such as /b/) have energy over a wide-range of frequencies (i.e., they are diffuse), but the energy is more concentrated in the lower frequency ranges. Alveolars (/d/) also have energy at a wide-range of frequencies, but have greater energy at the higher frequencies. Velars (such as /g/) have a concentration of energy in the middle-frequency range, and less energy at lower or higher frequencies. This is shown in Figures 8-10, which have spectrums for /bæ/, /dæ/ and /gæ/ respectively. Here, frequency is on the x-axis, and amplitude on the y-axis. The /b/ spectrum has a downward slope, or more energy at low frequencies. The /d/ spectrum has more energy at the high frequencies than does the /b/, and /g/ has most of its energy in the center of


80
60
40
20
0
0
1
2
3
4
5
frequency (kHz)


80
60
40
20
0
0
1
2
3
4
5
frequency (kHz)


80
60
40
20
0
0
1
2
3
4
5
frequency (kHz)

the spectrum. Stevens and Blumstein found that templates based on these verbal descriptions were quite accurate at classifying stops in syllable-initial position (averaging 85% correct acceptance by template), but were not as accurate in final position (approximately 76% correct acceptance; Blumstein & Stevens, 1979).

Unfortunately, follow-up research was not as positive. Walley and Carrell (1983) showed that when spectral tilt and formant frequency values were placed in opposition, listeners identified the phoneme according to the formant frequencies. Blumstein, Isaacs and Mertus (1982) also showed that when stimuli had onset spectra that conflicted with their formant frequencies, listeners' responses were dominated by the formant frequencies. However, in both experiments classification performance deteriorated when the information conflicted, suggesting that onset spectra were still used as a cue by listeners, even if it was not the primary one.

Kewley-Port (1983; Kewley-Port, Pisoni & Studdert-Kennedy, 1983; Kewley-Port & Luce, 1984) suggested that a dynamic measure of spectral-tilt over time would be better at classifying phonemes. She suggested three time-varying features which could be used to classify place of articulation for initial stops: the tilt of the spectrum at onset (rising vs. falling); the occurrence of high-amplitude, low-frequency energy late in the spectrum; and the presence of a single, prominent mid-frequency peak extending over time. Human observers could use these cues to classify phonemes correctly 88% of the time (Kewley-Port, 1983). Furthermore, listeners classified stops better when presented with just these dynamic cues than when presented with just the static spectral properties (Kewley-Port et al., 1983).

However, Lahiri, Gewirth and Blumstein (1984) found that even these changes were not sufficient. Although they were appropriate for English consonants, they were not capable of distinguishing labial from dental stops, even though some languages make this distinction. Furthermore, they did not classify dentals and alveolars as being the same place of articulation, even though linguistic theory labels them both as coronals. The authors suggested that measuring spectral-tilt at two different points in time (stop release and voicing onset) and calculating the change between these two points was a better metric. That is, the changes in distribution of energy over time seemed to better classify stops across different languages. This has remained the latest version of the theory.

Lahiri *et al.*'s metric, like those of Stevens and Blumstein (1979), required a human observer to classify the phonemes. Sawusch (1988) developed a computational version of this metric. This will be the version examined in this experiment. However, spectral tilt is highly correlated with spectral moments (described below). It is unlikely that spectral tilt would demonstrate strong perception-production links if spectral moments do not also do so. For that reason, spectral moments will be examined first, and tilt will only be examined if the moments data suggest there is something present worth investigating.

Spectral moments. Forrest, Weismer, Milenkovic and Dougall (1988) suggested that word-initial voiceless obstruents (stops and fricatives) could be identified from their spectrum by computing the spectral moments. The mean, variance, skewness and kurtosis of the noise portion at onset would summarize the concentration, the tilt, and the peakedness of the energy distribution (the same characteristics that the spectral tilt metric was trying to capture). The fricative centroid of Experiment 2 is the same as the mean, here. A cross-section of the spectrum is examined, and the distribution of energy across different frequencies is tabulated. From this distribution, the mean value, variance, skewness, and kurtosis (similar to the diffuse/compact distinction of Stevens and Blumstein) can be calculated. Forrest *et al.* found that this combination of features did distinguish between the places of articulation for stop consonants. For example, /p/ and /t/ differ from one another in skewness and mean, whereas /k/ differs from both of these in kurtosis. A linear discriminant analysis calculated from the first 10 ms of the three voiceless stops correctly classified them approximately 80% of the time. Calculating the moments from the first 40 ms of the signal improved classification to 92% accuracy. The fricatives

/s/ and /ʃ/ were classified even more accurately, although the moments failed to discriminate between /f/ and /θ/.

Tomiak (1991) examined this metric in further detail. She found that it was capable of classifying 74-78% of clear tokens of all voiceless fricatives (/s, ʃ, h, f, θ/), and an average of 92% of tokens of /s/, /ʃ/, and /h/ alone. Furthermore, when peak and moment information conflicted, listeners showed a tendency to classify phonemes according to the moment information, suggesting that this metric may indeed be related to cues listeners actually use. On the other hand, classification of even high-quality, well-identified stimuli was far poorer than human judgments, leaving these conclusions somewhat in doubt.

Richardson (1992) attempted to apply this metric to the 6 English stop consonants (both voiceless and voiced), and found much poorer classification. Performance averaged only 50% correct. He suggested that this metric may play some role in human classification (since performance was far greater than a chance score of 17%) but is unlikely to be a sufficient cue.

Sawusch and Dutton (1992) also attempted to apply this metric to stops. They found 88% classification for the three voiced stops. This average is substantially better than that found by Richardson (1992.) However, Richardson examined 1,385 tokens, whereas Sawusch and Dutton examined only 48 (in addition to the fact that he had examined all 6 English stop consonants, rather than just the 3 voiced ones). Thus, Sawusch and Dutton's stop consonants likely had far less variability among tokens within the same category, which would serve to increase the percentage of correct classification. Sawusch and Dutton also applied the metric to vowels, although they only attempted to determine whether there were unique prototypical patterns for the different vowels, rather than attempting to classify them. Although they did find some dimensions that seemed to correlate with vowel features (higher means for front vowels, higher kurtosis for tense vowels), the variability was also quite high, suggesting that it would be difficult to use this metric to classify vowels.

Peak differences. Syrdal and Gopal (1986) suggested that the frequency differences between formants might be a useful cue for classifying vowels. (Fischer-Jørgensen had made a similar proposal much earlier, but had not followed up on it; see Fischer-Jørgensen, 1954.) That is, although the exact values of formants may vary across individuals, their relative locations are more consistent. (Since formants are those frequencies emphasized by the shape of the oral cavity, individuals with different-shaped vocal tracts will have different formant values even when producing the same sound. However, these inter-talker differences in vocal tract morphology are likely to affect all the formants to a similar degree.

Thus, subtracting one formant from another should serve to normalize the signal for these talker differences.) More specifically, Syrdal and Gopal transformed the fundamental and formant frequencies to a critical band (or Bark) scale, which is believed to be a better approximation of the scaling functions of the human peripheral auditory system. Then, they calculated Bark-difference scores for F1-F0, F2-F1, F3-F2, F4-F3, and F4-F2. Vowels were classified on the basis of whether these differences were larger or smaller than a critical distance of 3 Barks. This critical distance was suggested in prior work by Chistovich and colleagues (Chistovich, Sheikin & Lublinskaja, 1979; Chistovich & Lublinskaya, 1979). Syrdal and Gopal found that the F1-F0 difference is related to how high a vowel is: High vowels have a Bark-distance less than the critical distance, while mid and low vowels have a Bark-distance greater than 3 Barks. F3-F2 is related to how front a vowel is, with back vowels exceeding the critical distance, but not front vowels. Thus, the authors suggest that these differences may be used to classify vowels across many different talkers.

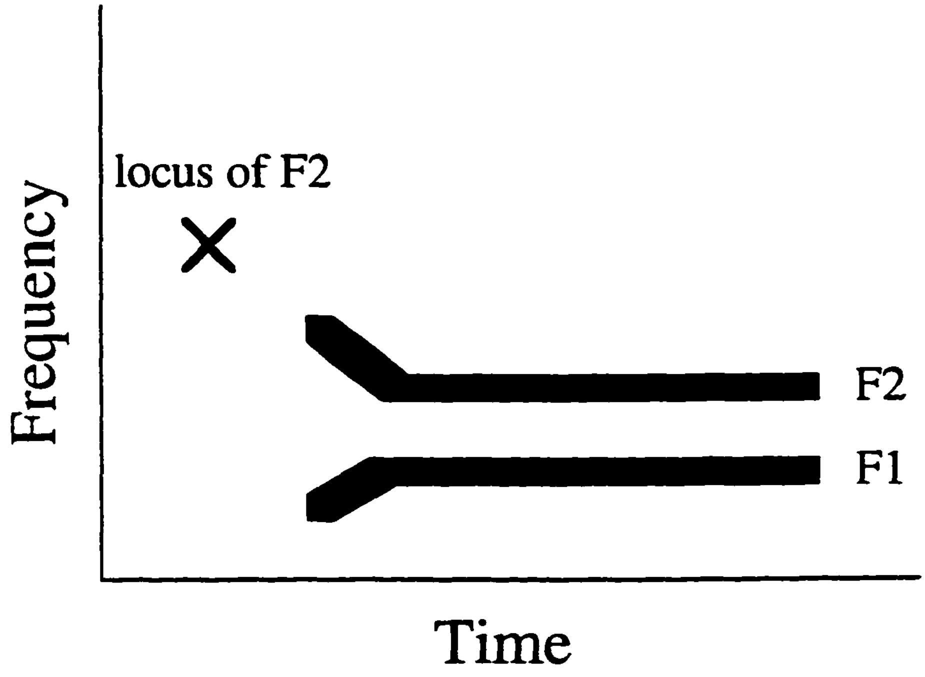
Sawusch and Dutton (1992) followed up on this idea, and developed a metric on this basis which could be used on all phonemes (rather than just vowels). Instead of basing decisions on binary features (< 3 Barks vs. > 3 Barks), as did Syrdal and Gopal (1986), they found prototypical values for each phoneme on all five difference scores, and classified new items according to the most similar prototype. Unfortunately, this classification scheme did not work well for high vowels. The authors then attempted to use this metric on voiced stop consonants, and found 88% correct classification.

Richardson (1992) also attempted to evaluate peak differences on the classification of stop consonants. He used both voiced and voiceless stops, and (as with his results with spectral moments) found that classification performance was quite poor overall (averaging 37% correct for static peak differences, and 35% for dynamic peak differences, across all six stops), although still above chance. As with spectral moments, Richardson found a much lower percentage correct than did Sawusch and Dutton. This is likely due to the fact that he examined many more tokens than did the other researchers, thus capturing the model's performance in a high-variability situation. He suggests that peak differences (like moments) may be used by human listeners, but are not sufficient by themselves.

Although the classification results from these more recent studies are not especially encouraging, the high classification for stops found by Sawusch and Dutton (1992) leave some room for hope. While Richardson (1992) is likely correct that this cue cannot be sufficient by itself, it may still be one of a set of cues used by listeners.

Locus equations. The idea that the locus (or starting point) of a formant transition could be used to differentiate places of articulation was first suggested by Delattre, Liberman and Cooper (1955). They suggested that the locus of F2 was important for place of articulation in stop consonants (and possibly in other consonants as well). More specifically, they suggested that /b/ has a locus of 720 Hz, /d/'s locus is 1800 Hz, and that /g/ has a 3000 Hz locus for front vowels but no locus for back vowels. (These loci are not the actual frequency of the formant transition at onset, but are rather what one would find if the formant were extrapolated back prior to the onset, or the location "to which [the formant] may be assumed to 'point' " (Delattre et al., 1955 p. 769). The locus might be thought to represent the idealized starting point of the consonant, and thus indicates the configuration of the articulators at the consonant's theoretical starting point.) An example of an F2 frequency locus is shown in Figure 11.

Lindblom (1963) suggested that by measuring F2 at onset and at midvowel, and making straight line regression fits between these two points for a number of CV tokens, it is possible to come up with equations that specify the coarticulation between the consonant and the vowel. He found that these "locus equations" had different slopes for different places of articulation, and thus could be used as a means of classifying phonemes.


Frequency
locus of F2
F2
F1
Time

Sussman and his colleagues (Sussman et al., 1991; Sussman et al., 1993; Sussman, 1991; Sussman, 1989) have followed up on this research, and suggested that these locus equations could be used to recover stop consonant place of articulation. They also have suggested a metric by which these equations could be calculated by the auditory system. Their algorithm was relatively successful, and a discriminant analysis classified the consonants correctly 83% of the time, if the velar stops in a back vowel context were not included (these had much poorer classification, see Sussman et al., 1991). Furthermore, these locus equations may not be specific to English. Sussman, Hoemeke and Ahmed (1993) found locus equations for stops in Thai, Arabic and Urdu, and found a high correlation for the locus equations in the different languages. This suggests that these cues may be tapping something related to an abstract notion of place of articulation.

Fowler (1994), on the other hand, has argued that locus equations really provide a measure of coarticulation, and only provide information for place of articulation indirectly. As such, they would also be affected by differences in manner of articulation. That is, the loci may be able to distinguish consonants when place of articulation is the only feature that is varying, but would not be able to do so when there was other information (such as manner) changing as well. Further, she found that the locus equations for /d/ and /z/ were significantly different from one another, even though they are produced with the same place of articulation. This suggests that locus equations do not provide invariant information for place of articulation (although this may not be relevant to the cues' usefulness for distinguishing stops). Perhaps more problematic, she found that while the locus equations for average productions of /b/ and /d/ differed, any given production might not fall closest to its own regression line. That is, the mean values for /b/ and /d/ were distinct, but there was sufficient overlap to make the locus equations a poor method of discrimination. In fact, Fowler found only 70% correct classification of /b, d, g/ for males, and 62.5% for females.

These results suggest that locus equations may not be as good a method of classifying consonants as Sussman's research has suggested. Nevertheless, it may still be related to a cue used by listeners, even if it is unlikely to be the only such cue.

Contrasting metrics

Unfortunately, it is impossible to create speech series that contrast all of these metrics. The metrics do not refer to completely different information in the spectrum, but instead refer to different ways of describing the same information. While there have been attempts to contrast some of these metrics (Sawusch & Dutton, 1992; Richardson, 1992; Tomiak, 1991), others are too closely related for this to be possible. For instance, the Peak Difference Metric and the Locus Metric both are based (at least in part) on the location of F2. Changing F2 necessarily changes both metrics, and this makes it difficult to contrast these metrics experimentally.

In the present experiment, a different way of evaluating these metrics is proposed. If the degree of perception-production correlation on a given cue is based on the extent that cue is related to the perceptual dimensions the listener actually uses, then the degree of correlation can be used as means of evaluating this relation. Thus, this methodology allows for a way of assessing the relative usefulness of these metrics. Whichever metric results in the greatest perception-production correlation would be suggested to be the metric most related to what humans actually use. This makes the assumption that perception and production are in fact linked, and that the degree of correlation between the two modalities depends on the appropriateness of the cue being measured. The results from Experiment 1 provide some support for this hypothesis. However, if further research throws these results into question, the results from the current experiment would necessarily be thrown into question as well.

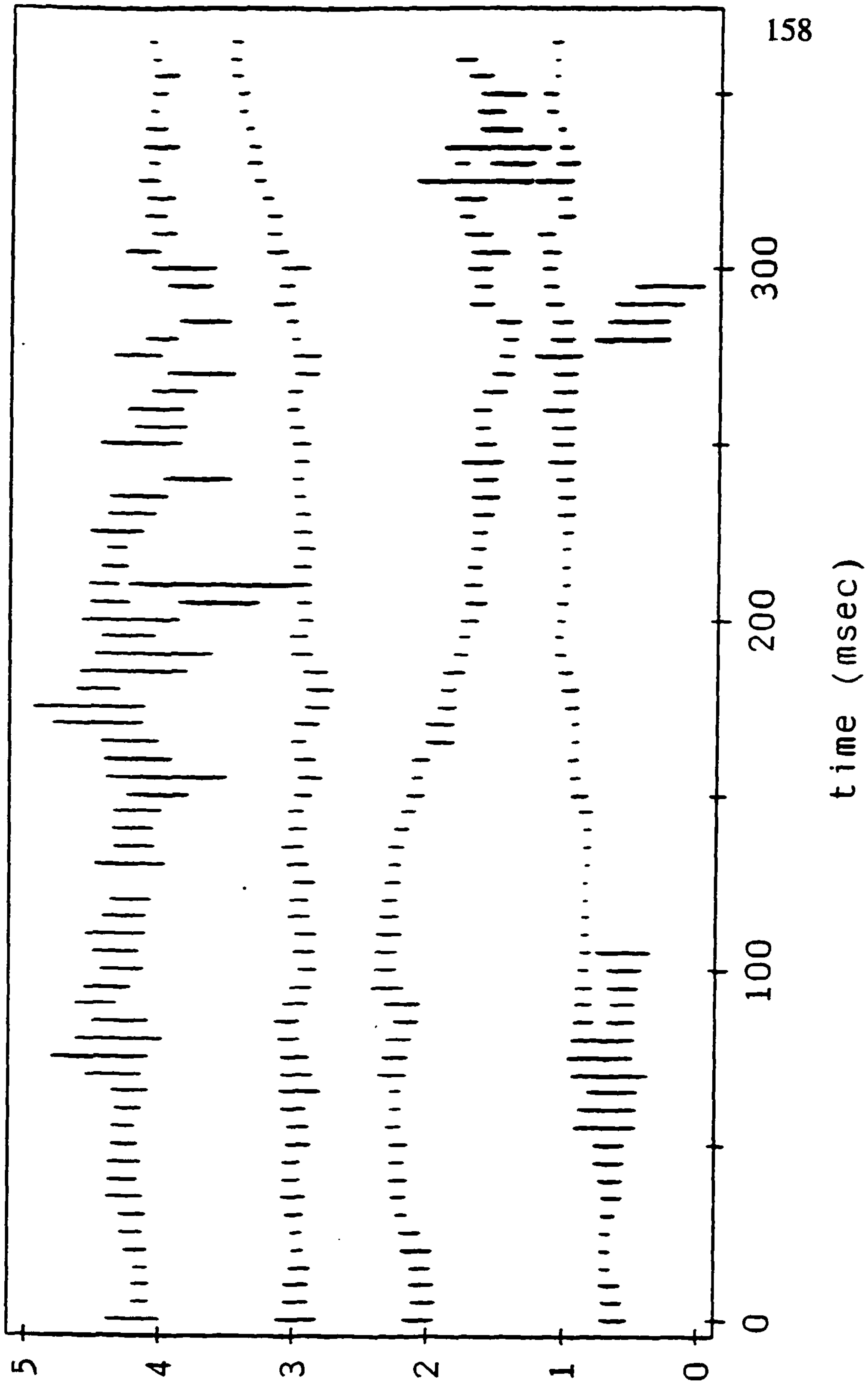
While this methodology (examining various metrics to see which produces the greatest perception-production correlation) works in theory, in practice there is some risk of spuriously high correlations, especially when only one target phoneme is being examined. For this reason, it is better to examine a number of phonemes with each metric, and to look for the pattern of correlations across these phones. If one metric has a larger perception-production correlation than the others on a variety of different phonetic prototypes, it would strongly suggest that that metric is more closely related to the cues listeners are actually using, and thus is perhaps a more promising metric for future study.

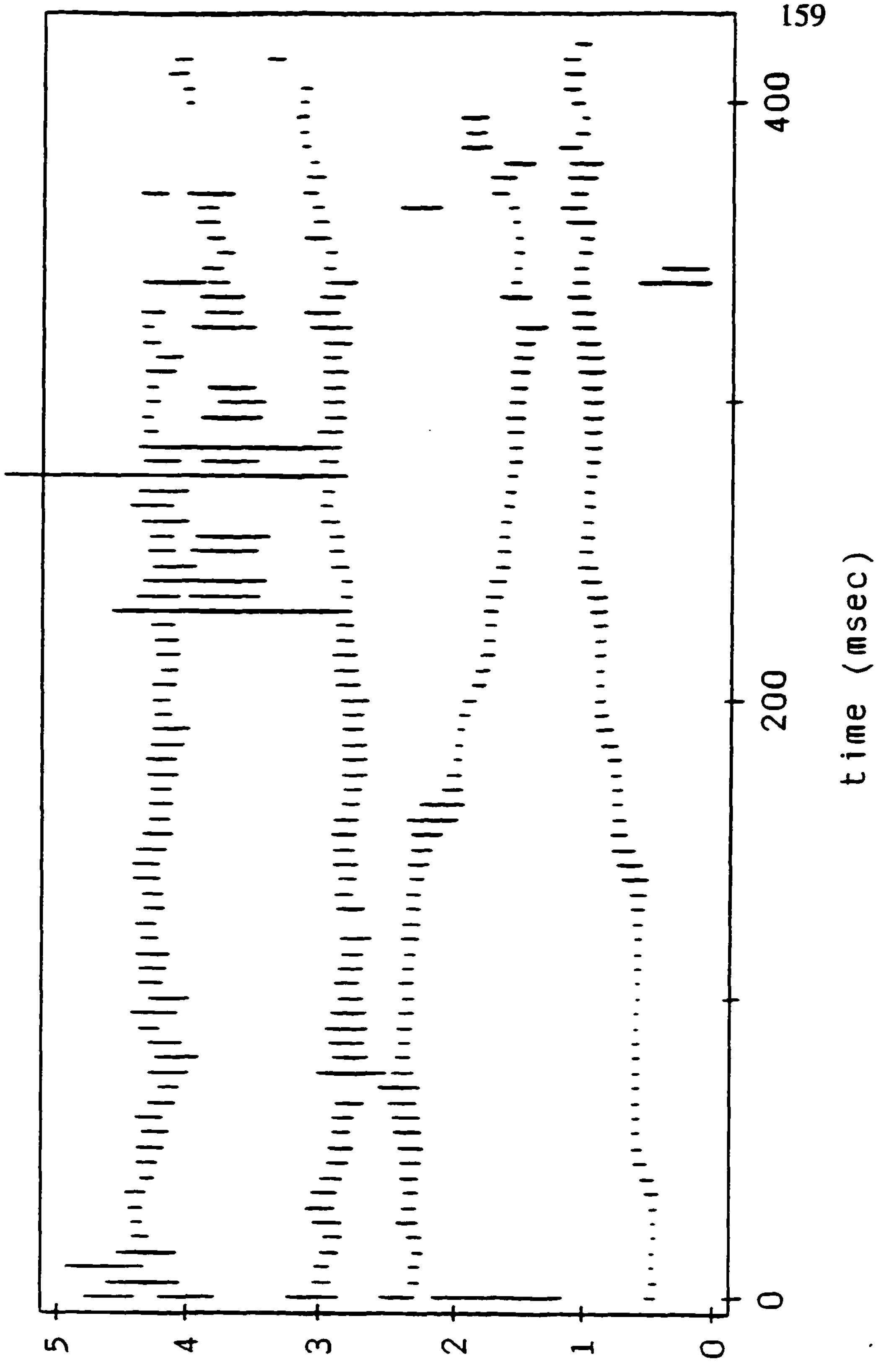
The present experiment attempts to do just this. However, in order to make the experiment feasible from a practical standpoint, some procedural changes need to be made. Because these prototype experiments require a fair amount of time from each subject, to actually test each metric individually on a number of different phonemes would not be possible, at least not in a within-subjects design (assuming it would be possible to experimentally contrast the different metrics, which has already been noted to be a problem). Furthermore, because these metrics are all based on combinations of cues, and the cues in different metrics are often related, it is not possible to make series whose endpoints only differ in phonetic category according to one metric. That is, one cannot make a series of items which differ according to the spectral moments metric without also having them differ to some extent in the other metrics as well, especially if one does not wish to use degraded speech (such as 2-formant stops). An additional problem is that, unlike the first two experiments, in which there was a single cue that could differentiate the two phonemes (VOT for /p/ vs. /b/, frication centroid for /s/ vs. /ʃ/), there are many sets of phonemes for which a single distinctive cue cannot be found. So, it is not possible to individually manipulate a single cue for each metric, and to use this as a way of finding the perception-production correlations.

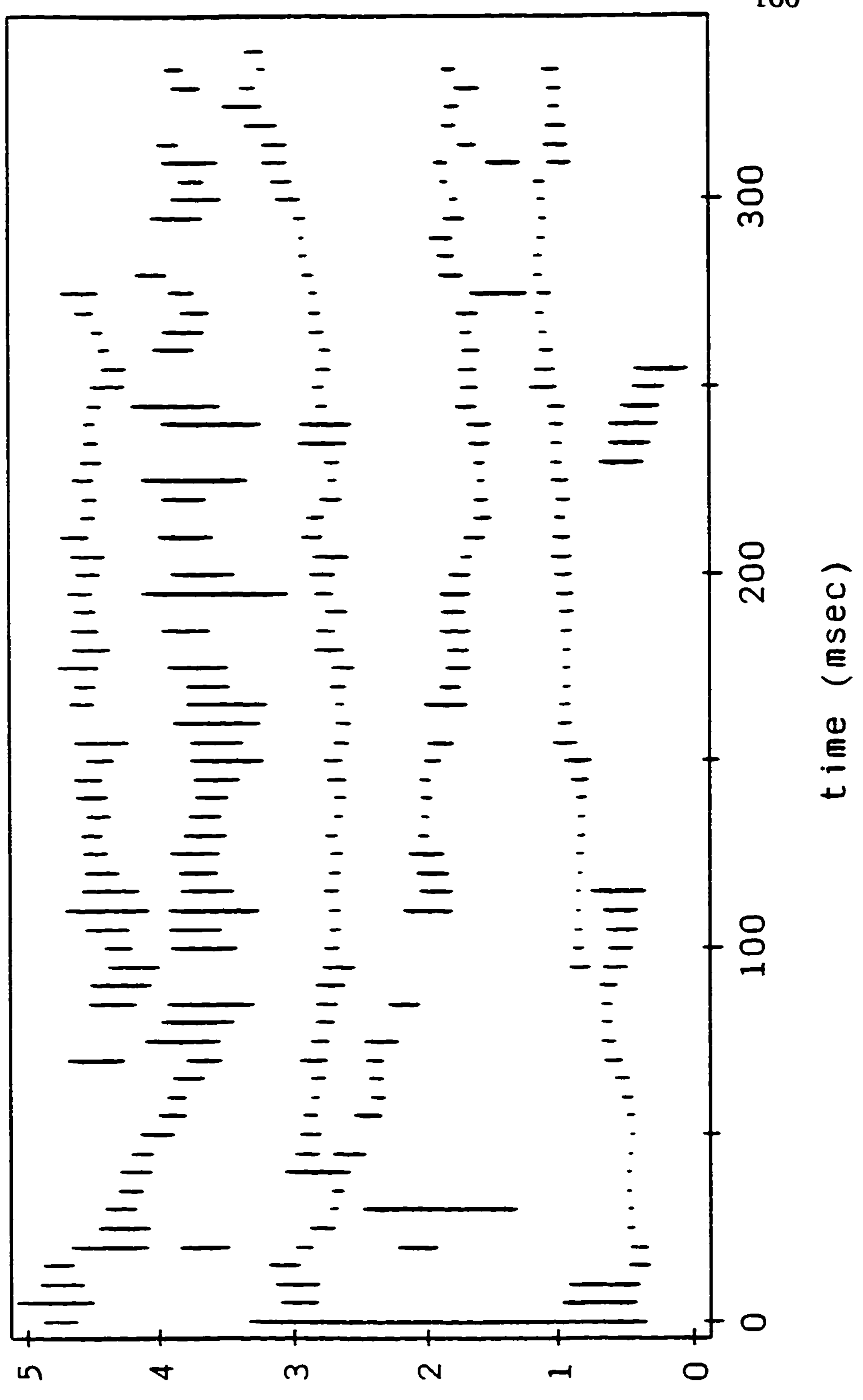
To get around these difficulties, this experiment uses series in which multiple cues are varying at one time, in a manner similar to that in natural speech. Natural tokens will be selected from several phonemic categories, and frequency values will be interpolated between them to make several continua. Subjects will listen to only a single series for a target phoneme, and their prototypes will be determined in a manner similar to those of the prior experiments. Then, the values on each metric will be calculated for that individual's perceptual prototype and production tokens. Separate multiple regressions will be run for each metric, even though the data points are being measured on the same perceptual series.

This substantially reduces the number of hours required from each subject in order to perform the experiment, making it feasible. However, it still requires that each subject perform the perceptual experiment multiple times, once for each phonemic target. Since the experiment has taken two to three days to run for a single phoneme, this would still result in at least six hours of subject time being necessary in order to find results from three phonemes. To further reduce this time requirement, phonemes were chosen which are bounded on both sides by other phonemes. In the first experiment, /p/ was bounded on one side by /b/, but was not bounded on the other side. This bounding forces the ratings to drop off at a faster rate, although it should not alter the existence of a prototype. (Thus, ratings dropped off faster towards the /ba/ end of the series in Experiment 1 than they did towards the /*pɑ/ end.) In Experiment 2, the target phoneme was also bounded on only one side: the target /ʃ/ will sound more and more like /s/ as the frequency centroid increases, but will not become more like any other phoneme in English when the centroid decreases. Since the ratings drop off faster when the prototype is bounded, there need not be quite as many stimulus items presented for each target phone if the target is bounded on both sides rather than just one.

Here, the target items were /b/, /d/, and /g/. These three phonemes are produced by forming a closure in the mouth, and then releasing it after pressure has built up. They differ in the location of that occlusion, or in their place of articulation. The velar consonant, /g/, is produced the furthest back in the mouth, the /b/ is produced furthest forward, and the /d/ is intermediate. Figures 12-14 show the formant patterns for tokens of


0
1
2
3
4
5
0
100
200
300
time (msec)


5
4
3
2
1
0
0
200
400
time (msec)


time (msec)
0
100
200
300
0
1
2
3
4
5

/bæ/, /dæ/, and /gæ/ respectively. Here, time is on the x-axis, and frequency is on the y-axis. The differences in the formant patterns of these three consonants are primarily in the locations of the second and third formants at the beginning of the syllable. It is possible to make natural-sounding synthetic series ranging from /b/ to /d/, from /d/ to /g/, and from /g/ to /b/ by interpolating between the locations of the formants in natural productions of these syllables. In this manner, each of these three target phones are bounded on both sides by one of the alternatives, which will lessen the number of stimuli needed for presentation to subjects in order to get a good measure of a prototype. Another advantage of using these phones is that all of the metrics described above can be easily measured on them. This is not the case for all phonemes. For example, the peak difference metric could not be applied to the frication portion of a voiceless fricative, such as /s/, as there are no measurable formants in the noise. All of the metrics discussed have been applied to voiced stops in the literature, making these phones ideal choices.

One final change was made to the experimental procedure to further decrease the time constraints. Rather than presenting the stimuli in random order with a fixed number of presentations per stimuli, they were presented using an adaptive testing method. This type of presentation method is based on the classic method of limits. Rather than presenting all stimuli the same number of times in a random order, stimuli will be presented in an ascending/descending method. Stimuli from one extreme end of the series, which are expected to be rated poorly, will be presented first. Then, stimuli slightly further from this extreme will be presented. As long as the subjects' ratings increase, stimuli closer and closer to the opposite extreme will be presented. When ratings start decreasing once again, the selection of stimuli for presentation will reverse. In this manner, most presentations will occur in the region hovering around that individual's prototype. The items rated as poor examples will be presented fewer times to subjects than will the items rated relatively highly. Since the focus of the task is to determine the prototype, the poorly-rated items are not of interest, and this procedure should be much faster than the method of constant stimuli used in Experiments 1 and 2 (see Sawusch, 1996). This change in procedure allows for even shorter time requirements, without reducing sensitivity in the region which is of primary interest.

To summarize, the current experiment examines the perception-production relations for three consonants: /b/, /d/, and /g/. Listeners will be asked to rate tokens from each phoneme category, as well as to produce tokens from all three categories. Both the perceptual prototypes and the productions will be analyzed according to three or four metrics: spectral moments, peak differences, and F2 loci (and possibly spectral tilt, depending on the results of the spectral moments data). If one of these metrics is more closely related to the cues listeners actually use, there should be stronger perception-production correlations for that metric, across all three phonemes. If there are no differences in these correlations, or if the differences are not consistent across the three phonemes, it would not be possible to determine whether any of these metrics are more accurate ways of describing place-of-articulation information than are the other metrics.

Method

Subjects. Thirty-five subjects participated in this experiment, which required 3 one-hour sessions. Subjects received $15 in compensation at the end of the third day of the experiment (three subjects also received course credit). All subjects (with one exception) were native speakers of English, with no history of hearing disorders. One subject was found during questioning to be a native speaker of Spanish, rather than English; her data are not included. Six subjects reported having a second language spoken in their home (2 Spanish, 2 Chinese, 1 Korean, 1 French), although English was still their primary language. Data from these subjects were included in the analysis. One additional subject had had some articulation difficulties as a child (tongue thrust), but had normal production at the time of the experiment. All other subjects reported normal articulation. One subject failed to complete the experiment. Her data were not included. This left a total of 33 subjects.

Subjects were asked to complete a survey regarding their dialect-background before participating in this experiment. Most of the participants were born and grew up on the east coast. Of these, 9 were from New York City, 2 from Long Island, and the rest from other locations in New York or New Jersey. Approximately one-fourth of the subjects were not raised in the east: one subject was born and raised in Toronto, a second was born and raised in California, 2 others were born in the midwest (OH or MI) before moving to New York, and one spent a fair amount of his childhood in Florida.

Stimuli. For the production task, a female native talker of English (RSN) recorded six tokens of each CV syllable beginning with either /b/, /d/, or /g/, and followed by the vowel /æ/, and three tokens of each of the other CV syllables consisting of /b/, /d/, /g/, /p/, /t/ or /k/ and followed by the 7 vowels /i, e, æ, u, o, ɑ, ʌ/. All of the tokens were amplified, low-pass filtered at 9.5 kHz, digitized via a 16-bit, analog-to-digital converter at a 20 kHz sampling rate and stored on computer disk.

For the perception task, the stimuli were created synthetically, as there is no way to edit a natural continuum based on slight formant frequency differences. The stimuli were based on high-quality natural tokens of /bæ/, /dæ/, and /gæ/ from a male talker. A male talker was chosen because our synthesizer does a better job of mimicking male voices. These were synthesized, and used as endpoints. Values for the frequency, amplitude, and bandwidth for the first five formants, the fundamental frequency, and the amplitude of release burst frication and voicing were interpolated between each pair of endpoints in 20 equal steps. Three continua were made, one ranging from /b/ to /d/, one from /d/ to /g/, and the third from /g/ to /b/. Each continuum consisted of 21 items (including both endpoints). Thus, there were a total of 60 different syllables. The synthesis parameters for the three endpoints are shown in Tables 11-13.

Procedure. The procedure was similar to that used in Experiments 1 and 2. In the production task, subjects were asked to repeat each CV syllable they heard in their normal manner of production. In the perception task, subjects were asked to rate the stimuli as to how good of an example of /b/ they were in one session, as /d/ in a second session, and as /g/ in a third (each of the six possible orderings of these three sessions was presented to subjects in an alternating fashion). Subjects were not presented with all of the tokens in each session. When they were judging items as /d/, they heard only the items ranging from /b/ to /d/ and from /g/ to /d/, not those that range from /b/ to /g/, and likewise for other sessions.

Table 11 /bæ/ synthesis parameters

GLobal Parameters:

F Glt Res	B Glt Res	F Glt Zero	B Glt Zero	B Glt Res2
0	100	1500	6000	200
F6	B6	F Nsl Pol	B Nsl Pol	B Nsl Zero
5000	1000	250	100	100
Gain	Auto Amp	No.Cas For	C/P SW	Cor SW
36	-1	5	1	0

msec	FI	BI	AI	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
0	330	150	62	1300	500	52	2000	350	36	3200	250	28	4250	200	19	0	0	250	120	0	0	0	0	55
5	383	134	62	1470	450	52	2078	300	37	3266	240	30	4250	200	20	0	0	250	120	69	0	0	0	0
10	435	117	62	1491	420	53	2155	450	38	3242	229	33	4250	200	24	0	0	250	119	70	0	0	0	0
15	488	101	63	1513	330	53	2233	180	40	3218	224	36	4250	200	27	0	0	250	119	72	0	0	0	0
20	540	84	63	1534	220	54	2310	140	41	3194	200	39	4250	200	31	0	0	250	118	73	0	0	0	0
25	550	73	63	1546	193	54	2388	110	43	3170	150	41	4211	207	33	0	0	250	118	73	0	0	0	0
30	551	63	63	1557	165	55	2388	109	43	3173	160	42	4156	214	35	0	0	250	117	73	0	0	0	0
35	551	62	64	1569	138	56	2388	109	44	3162	170	43	4074	221	35	0	0	250	117	73	0	0	0	0
40	552	61	64	1580	110	57	2388	112	45	3156	171	44	4051	228	36	0	0	250	116	73	0	0	0	0
45	552	60	64	1592	108	58	2387	115	45	3168	172	45	4056	235	36	0	0	250	116	73	0	0	0	0
50	553	59	64	1603	107	59	2387	118	46	3200	184	45	4084	242	37	0	0	250	115	73	0	0	0	0
55	553	58	65	1564	105	60	2387	121	46	3231	195	44	4112	257	37	0	0	250	115	73	0	0	0	0
60	554	59	65	1547	108	61	2387	124	47	3253	207	43	4113	272	37	0	0	250	114	73	0	0	0	0
65	557	60	65	1556	109	61	2385	127	47	3282	218	43	4114	287	38	0	0	250	114	73	0	0	0	0
70	563	61	65	1585	109	61	2384	130	48	3283	230	42	4114	302	38	0	0	250	113	73	0	0	0	0
75	568	66	64	1637	110	59	2383	133	48	3283	246	41	4115	310	39	0	0	250	113	73	0	0	0	0
80	572	72	64	1668	110	59	2381	100	49	3284	256	41	4096	313	39	0	0	250	113	73	0	0	0	0
85	574	75	64	1697	109	59	2380	122	50	3284	258	42	4076	313	39	0	0	250	112	73	0	0	0	0
90	575	76	65	1682	107	59	2378	143	51	3285	261	42	4056	313	40	0	0	250	111	73	0	0	0	0
95	576	75	64	1682	106	60	2377	146	52	3262	263	42	4060	312	40	0	0	250	110	73	0	0	0	0
100	576	75	64	1654	104	61	2376	149	53	3238	265	42	4063	312	40	0	0	250	110	73	0	0	0	0
105	579	74	64	1651	103	61	2374	152	53	3215	260	42	4067	312	40	0	0	250	109	73	0	0	0	0
110	583	73	64	1674	101	61	2373	146	53	3211	255	42	4071	312	40	0	0	250	108	73	0	0	0	0
115	586	72	64	1673	100	61	2371	140	53	3208	250	42	4074	311	40	0	0	250	107	73	0	0	0	0
120	585	71	63	1672	102	61	2370	135	53	3204	252	42	4078	311	40	0	0	250	107	73	0	0	0	0

Table 11, continued /bæ/ synthesis parameters

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
125	583	71	63	1671	104	60	2369	130	53	3200	255	42	4081	311	40	0	0	250	106	73	0	0	0	0
130	584	70	63	1670	106	60	2367	140	53	3183	257	42	4085	310	40	0	0	250	105	73	0	0	0	0
135	586	69	63	1669	108	60	2366	150	52	3166	260	42	4088	310	39	0	0	250	104	73	0	0	0	0
140	587	68	63	1668	109	60	2364	156	52	3148	262	42	4092	310	39	0	0	250	104	73	0	0	0	0
145	584	67	63	1654	111	60	2363	162	52	3131	263	42	4096	305	39	0	0	250	103	73	0	0	0	0
150	583	66	62	1657	113	60	2362	167	52	3114	264	42	4099	300	39	0	0	250	102	73	0	0	0	0
155	581	66	62	1653	115	60	2360	173	52	3097	265	42	4103	303	39	0	0	250	101	73	0	0	0	0
160	581	65	62	1636	117	60	2359	176	52	3091	266	42	4106	306	39	0	0	250	101	73	0	0	0	0
165	586	64	62	1619	117	59	2357	179	52	3084	267	42	4110	309	39	0	0	250	100	73	0	0	0	0
170	595	65	62	1602	118	59	2356	182	52	3077	268	42	4113	311	39	0	0	250	99	73	0	0	0	0
175	601	63	62	1586	116	59	2355	185	52	3071	269	42	4115	314	39	0	0	250	98	73	0	0	0	0
180	603	63	62	1569	116	59	2353	188	52	3064	270	42	4118	316	39	0	0	250	98	73	0	0	0	0
185	605	60	61	1552	117	59	2352	191	52	3057	278	42	4121	319	39	0	0	250	97	73	0	0	0	0
190	616	61	61	1535	124	59	2350	194	52	3050	281	42	4124	319	39	0	0	250	96	73	0	0	0	0
195	637	62	61	1536	130	59	2349	197	52	3044	278	42	4126	319	39	0	0	250	95	73	0	0	0	0
200	637	62	61	1537	136	59	2347	200	51	3037	274	42	4129	319	38	0	0	250	94	73	0	0	0	0
205	638	63	61	1537	143	58	2346	199	51	3033	271	42	4135	319	38	0	0	250	93	73	0	0	0	0
210	638	64	61	1538	150	58	2345	197	51	3030	268	42	4141	318	38	0	0	250	92	73	0	0	0	0
215	638	64	61	1539	157	58	2343	196	51	3026	265	42	4146	317	38	0	0	250	92	73	0	0	0	0
220	638	65	60	1523	164	58	2342	194	51	3023	260	42	4152	317	38	0	0	250	92	73	6	0	0	0
225	639	65	60	1520	164	58	2340	193	51	3019	257	42	4198	316	38	0	0	250	92	73	0	0	0	0
230	639	66	60	1517	172	58	2339	191	51	3016	255	42	4209	315	38	0	0	250	92	73	0	0	0	0
235	637	67	60	1514	180	58	2338	190	51	3012	255	42	4220	314	38	0	0	250	92	73	0	0	0	0
240	640	67	60	1511	184	58	2336	188	51	3009	255	42	4232	314	38	0	0	250	93	73	0	0	0	0
245	641	68	60	1507	188	57	2335	187	51	3005	255	42	4243	313	38	0	0	250	93	73	0	0	0	0
250	642	69	59	1504	193	57	2333	185	51	3001	255	42	4254	312	38	0	0	250	94	73	0	0	0	0
255	642	69	59	1501	198	57	2332	184	51	2998	255	42	4265	311	38	0	0	250	94	73	0	0	0	0
260	643	70	59	1498	202	57	2331	182	51	2994	255	42	4277	311	38	0	0	250	95	73	0	0	0	0
265	644	71	59	1498	207	57	2329	168	51	2991	255	42	4288	310	38	0	0	250	95	73	0	0	0	0
270	655	71	59	1498	211	57	2328	170	50	2987	255	42	4299	309	37	0	0	250	96	73	0	0	0	0
275	666	72	59	1497	216	57	2326	200	50	2984	255	42	4299	309	37	0	0	250	96	73	0	0	0	0
280	677	73	59	1495	221	57	2325	215	50	2980	255	42	4299	309	37	0	0	250	97	73	0	0	0	0
285	689	73	58	1494	225	56	2324	222	50	2977	255	42	4299	309	37	0	0	250	98	73	0	0	0	0
290	700	74	58	1492	230	56	2322	221	50	2973	257	42	4299	309	37	0	0	250	99	73	0	0	0	0
295	711	74	58	1491	235	56	2321	225	50	2969	264	42	4318	302	37	0	0	250	101	73	0	0	0	0
300	722	75	58	1493	239	56	2319	228	50	2966	271	42	4337	288	37	0	0	250	102	73	0	0	0	0
305	733	76	54	1495	244	53	2318	230	47	2962	278	39	4357	277	35	0	0	250	103	73	0	0	0	0

Table 11, continued /bæ/ synthesis parameters

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
310	744	76	51	1497	248	50	2317	230	43	2959	285	35	4376	257	32	0	0	250	104	73	0	0	0	0
315	755	77	47	1499	253	47	2315	228	40	2955	291	32	4395	240	30	0	0	250	105	73	0	0	0	0
320	767	78	44	1497	258	45	2314	226	36	2952	298	29	4414	227	27	0	0	250	106	73	0	0	0	0
325	778	78	40	1490	262	42	2312	224	33	2948	305	26	4434	219	25	0	0	250	108	73	0	0	0	0
330	789	79	37	1482	267	39	2311	222	29	2945	312	22	4453	210	22	0	0	250	109	73	0	0	0	0
335	800	79	33	1475	269	36	2310	221	26	2941	319	19	4472	206	20	0	0	250	110	73	0	0	0	0

Table 12 /dæ/ synthesis parameters

Global Parameters:

F Glt Res	B Glt Res	F Glt Zero	B Glt Zero	B Glt Res2
0	100	1500	6000	200
F6	**B6**	**F Nsl Pol**	**B Nsl Pol**	**B Nsl Zero**
5000	1000	250	100	100
Gain	**Auto Amp**	**No.Cas For**	**C/P SW**	**Cor SW**
36	-1	5	1	0

msec	Fl	Bl	Al	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
0	350	250	35	1680	200	37	2700	350	35	3533	300	36	4000	450	34	0	0	250	118	0	0	0	0	77
5	377	236	42	1698	250	39	2650	200	37	3561	290	37	4012	405	35	0	0	250	118	0	0	0	0	54
10	404	221	48	1715	300	41	2600	280	39	3552	280	38	4025	379	36	0	0	250	118	69	0	0	0	0
15	430	207	52	1733	286	44	2559	315	41	3459	240	39	4037	353	37	0	0	250	118	70	0	0	0	0
20	457	192	56	1750	272	46	2518	333	43	3366	200	40	4049	326	38	0	0	250	118	72	0	0	0	0
25	484	150	60	1711	240	48	2477	350	45	3334	200	41	4055	300	40	0	0	250	118	73	0	0	0	0
30	468	137	61	1705	212	50	2436	250	45	3299	200	37	4060	311	41	0	0	250	118	73	0	0	0	0
35	482	125	62	1699	165	55	2395	150	45	3294	200	38	4066	322	39	0	0	250	118	73	0	0	0	0
40	500	112	63	1693	119	57	2394	150	46	3289	211	38	4072	334	39	0	0	250	118	73	0	0	0	0
45	502	90	63	1686	121	57	2392	150	46	3283	222	39	4077	345	38	0	0	250	118	73	0	0	0	0
50	510	68	61	1680	123	56	2391	150	46	3278	234	39	4083	356	38	0	0	250	117	73	0	0	0	0
55	536	69	61	1674	124	54	2389	150	46	3273	245	40	4089	356	37	0	0	250	116	73	0	0	0	0
60	550	69	63	1668	126	56	23a8	150	47	3276	256	41	4094	354	38	0	0	250	115	73	0	0	0	0
65	553	65	62	1634	124	60	2386	150	47	3280	267	41	4100	351	39	0	0	250	115	73	0	0	0	0
70	556	70	63	1647	123	59	2385	150	47	3283	245	41	4135	349	39	0	0	250	114	73	0	0	0	0
75	568	66	64	1637	110	59	2383	133	48	3283	246	41	4115	310	39	0	0	250	113	73	0	0	0	0
80	572	72	64	1668	110	59	2381	100	49	3284	256	41	4096	313	39	0	0	250	113	73	0	0	0	0
85	574	75	64	1697	109	59	2380	122	50	3284	258	42	4076	313	39	0	0	250	112	73	0	0	0	0
90	575	76	65	1682	107	59	2378	143	51	3285	261	42	4056	313	40	0	0	250	111	73	0	0	0	0
95	576	75	64	1682	106	60	2377	146	52	3262	263	42	4060	312	40	0	0	250	110	73	0	0	0	0
100	576	75	64	1654	104	61	2376	149	53	3238	265	42	4063	312	40	0	0	250	110	73	0	0	0	0
105	579	74	64	1651	103	61	2374	152	53	3215	260	42	4067	312	40	0	0	250	109	73	0	0	0	0
110	583	73	64	1674	101	61	2373	146	53	3211	255	42	4071	312	40	0	0	250	108	73	0	0	0	0
115	586	72	64	1673	100	61	2371	140	53	3208	250	42	4074	311	40	0	0	250	107	73	0	0	0	0
120	585	71	63	1672	102	61	2370	135	53	3204	252	42	4078	311	40	0	0	250	107	73	0	0	0	0

Table 12, continued /dæ/ synthesis parameters

msec	Fl	Bl	Al	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
125	583	71	63	1671	104	60	2369	130	53	3200	255	42	4081	311	40	0	0	250	106	73	0	0	0	0
130	584	70	63	1670	106	60	2367	140	53	3183	257	42	4085	310	40	0	0	250	105	73	0	0	0	0
135	586	69	63	1669	108	60	2366	150	52	3166	260	42	4088	310	39	0	0	250	104	73	0	0	0	0
140	587	68	63	1668	109	60	2364	156	52	3148	262	42	4092	310	39	0	0	250	104	73	0	0	0	0
145	584	67	63	1654	111	60	2363	162	52	3131	263	42	4096	305	39	0	0	250	103	73	0	0	0	0
150	583	66	62	1657	113	60	2362	167	52	3114	264	42	4099	300	39	0	0	250	102	73	0	0	0	0
155	581	66	62	1653	115	60	2360	173	52	3097	265	42	4103	303	39	0	0	250	101	73	0	0	0	0
160	581	65	62	1636	117	60	2359	176	52	3091	266	42	4106	306	39	0	0	250	101	73	0	0	0	0
165	586	64	62	1619	117	59	2357	179	52	3084	267	42	4110	309	39	0	0	250	100	73	0	0	0	0
170	595	65	62	1602	118	59	2356	182	52	3077	268	42	4113	311	39	0	0	250	99	73	0	0	0	0
175	601	63	62	1586	116	59	2355	185	52	3071	269	42	4115	314	39	0	0	250	98	73	0	0	0	0
180	603	63	62	1569	116	59	2353	188	52	3064	270	42	4118	316	39	0	0	250	98	73	0	0	0	0
185	605	60	61	1552	117	59	2352	191	52	3057	278	42	4121	319	39	0	0	250	97	73	0	0	0	0
190	616	61	61	1535	124	59	2350	194	52	3050	281	42	4124	319	39	0	0	250	96	73	0	0	0	0
195	637	62	61	1536	130	59	2349	197	52	3044	278	42	4126	319	39	0	0	250	95	73	0	0	0	0
200	637	62	61	1537	136	59	2347	200	51	3037	274	42	4129	319	38	0	0	250	94	73	0	0	0	0
205	638	63	61	1537	143	58	2346	199	51	3033	271	42	4135	319	38	0	0	250	93	73	0	0	0	0
210	638	64	61	1538	150	58	2345	197	51	3030	268	42	4141	318	38	0	0	250	92	73	0	0	0	0
215	638	64	61	1539	157	58	2343	196	51	3026	265	42	4146	317	38	0	0	250	92	73	0	0	0	0
220	638	65	60	1523	164	58	2342	194	51	3023	260	42	4152	317	38	0	0	250	92	73	6	0	0	0
225	639	65	60	1520	164	58	2340	193	51	3019	257	42	4198	316	38	0	0	250	92	73	0	0	0	0
230	639	66	60	1517	172	58	2339	191	51	3016	255	42	4209	315	38	0	0	250	92	73	0	0	0	0
235	637	67	60	1514	180	58	2338	190	51	3012	255	42	4220	314	38	0	0	250	92	73	0	0	0	0
240	640	67	60	1511	184	58	2336	188	51	3009	255	42	4232	314	38	0	0	250	93	73	0	0	0	0
245	641	68	60	1507	188	57	2335	187	51	3005	255	42	4243	313	38	0	0	250	93	73	0	0	0	0
250	642	69	59	1504	193	57	2333	185	51	3001	255	42	4254	312	38	0	0	250	94	73	0	0	0	0
255	642	69	59	1501	198	57	2332	184	51	2998	255	42	4265	311	38	0	0	250	94	73	0	0	0	0
260	643	70	59	1498	202	57	2331	182	51	2994	255	42	4277	311	38	0	0	250	95	73	0	0	0	0
265	644	71	59	1498	207	57	2329	168	51	2991	255	42	4288	310	38	0	0	250	95	73	0	0	0	0
270	655	71	59	1498	211	57	2328	170	50	2987	255	42	4299	309	37	0	0	250	96	73	0	0	0	0
275	666	72	59	1497	216	57	2326	200	50	2984	255	42	4299	309	37	0	0	250	96	73	0	0	0	0
280	677	73	59	1495	221	57	2325	215	50	2980	255	42	4299	309	37	0	0	250	97	73	0	0	0	0
285	689	73	58	1494	225	56	2324	222	50	2977	255	42	4299	309	37	0	0	250	98	73	0	0	0	0
290	700	74	58	1492	230	56	2322	221	50	2973	257	42	4299	309	37	0	0	250	99	73	0	0	0	0
295	711	74	58	1491	235	56	2321	225	50	2969	264	42	4318	302	37	0	0	250	101	73	0	0	0	0
300	722	75	58	1493	239	56	2319	228	50	2966	271	42	4337	288	37	0	0	250	102	73	0	0	0	0
305	733	76	54	1495	244	53	2318	230	47	2962	278	39	4357	277	35	0	0	250	103	73	0	0	0	0

Table 12, continued /dæ/ synthesis parameters

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
310	744	76	51	1497	248	50	2317	230	43	2959	285	35	4376	257	32	0	0	250	104	73	0	0	0	0
315	755	77	47	1499	253	47	2315	228	40	2955	291	32	4395	240	30	0	0	250	105	73	0	0	0	0
320	767	78	44	1497	258	45	2314	226	36	2952	298	29	4414	227	27	0	0	250	106	73	0	0	0	0
325	778	78	40	1490	262	42	2312	224	33	2948	305	26	4434	219	25	0	0	250	108	73	0	0	0	0
330	789	79	37	1482	267	39	2311	222	29	2945	312	22	4453	210	22	0	0	250	109	73	0	0	0	0
335	800	79	33	1475	269	36	2310	221	26	2941	319	19	4472	206	20	0	0	250	110	73	0	0	0	0

Table 13 /gæ/ synthesis parameters

GLobal Parameters:

F Glt Res	B Glt Res	F Glt Zero	B Glt Zero	B Glt Res2
0	100	1500	6000	200
F6	B6	F Nsl Pol	B Nsl Pol	B Nsl Zero
5000	1000	250	100	100
Gain	Auto Amp	No.Cas For	C/P SW	Cor SW
36	-1	5	1	0

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
0	360	130	59	2100	250	35	2300	82	38	3390	400	18	4500	400	10	0	0	250	120	0	0	0	0	71
5	360	122	59	2115	227	38	2329	166	40	3390	371	22	4488	366	15	0	0	250	120	0	0	0	0	74
10	360	114	57	2129	204	41	2358	171	41	3390	341	27	4476	332	20	0	0	250	120	0	0	0	0	74
15	360	106	56	2144	181	44	2388	175	43	3390	312	31	4464	308	25	0	0	250	120	0	0	0	0	54
20	360	111	56	2158	158	46	2417	162	44	3390	282	35	4452	401	30	0	0	250	120	69	0	0	0	0
25	386	116	56	2101	148	48	2446	148	46	3288	262	37	4440	228	35	0	0	250	120	70	0	0	0	0
30	413	121	55	2044	135	50	2475	143	48	3186	242	42	4429	200	35	0	0	250	120	72	0	0	0	0
35	439	126	60	1986	123	53	2451	138	47	3084	221	42	4417	99	36	0	0	250	120	73	0	0	0	0
40	465	162	62	1929	110	53	2428	133	48	3183	193	42	4405	217	36	0	0	250	120	73	0	0	0	0
45	491	120	62	1872	110	53	2404	135	48	3281	287	42	4393	217	36	0	0	250	120	73	0	0	0	0
50	518	199	62	1815	110	54	2380	128	49	3278	287	42	4381	217	36	0	0	250	120	73	0	0	0	0
55	544	157	62	1757	110	55	2380	109	49	3278	287	42	4369	385	37	0	0	250	119	73	0	0	0	0
60	570	141	62	1700	110	56	2381	124	49	3278	287	42	4328	421	37	0	0	250	118	73	0	0	0	0
65	564	125	61	1691	110	56	2381	126	48	3361	421	42	4288	437	37	0	0	250	116	73	0	0	0	0
70	559	109	64	1682	113	57	2382	128	48	3340	326	42	4247	457	37	0	0	250	115	73	0	0	0	0
75	553	93	64	1636	116	58	2382	130	48	3236	231	42	4108	196	38	0	0	250	115	73	0	0	0	0
80	553	84	65	1639	118	58	2383	132	48	3270	218	42	4141	273	39	0	0	250	114	73	0	0	0	0
85	568	66	64	1637	110	59	2383	133	48	3283	246	41	4115	310	39	0	0	250	113	73	0	0	0	0
90	572	72	64	1668	110	59	2381	100	49	3284	256	41	4096	313	39	0	0	250	113	73	0	0	0	0
95	574	75	64	1697	109	59	2380	122	50	3284	258	42	4076	313	39	0	0	250	112	73	0	0	0	0
100	575	76	65	1682	107	59	2378	143	51	3285	261	42	4056	313	40	0	0	250	111	73	0	0	0	0
105	576	75	64	1682	106	60	2377	146	52	3262	263	42	4060	312	40	0	0	250	110	73	0	0	0	0
110	576	75	64	1654	104	61	2376	149	53	3238	265	42	4063	312	40	0	0	250	110	73	0	0	0	0
115	579	74	64	1651	103	61	2374	152	53	3215	260	42	4067	312	40	0	0	250	109	73	0	0	0	0
120	583	73	64	1674	101	61	2373	146	53	3211	255	42	4071	312	40	0	0	250	108	73	0	0	0	0

Table 13, continued /gæ/ synthesis parameters

msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
125	586	72	64	1673	100	61	2371	140	53	3208	250	42	4074	311	40	0	0	250	107	73	0	0	0	0
130	585	71	63	1672	102	61	2370	135	53	3204	252	42	4078	311	40	0	0	250	107	73	0	0	0	0
135	583	71	63	1671	104	60	2369	130	53	3200	255	42	4081	311	40	0	0	250	106	73	0	0	0	0
140	584	70	63	1670	106	60	2367	140	53	3183	257	42	4085	310	40	0	0	250	105	73	0	0	0	0
145	586	69	63	1669	108	60	2366	150	52	3166	260	42	4088	310	39	0	0	250	104	73	0	0	0	0
150	587	68	63	1668	109	60	2364	156	52	3148	262	42	4092	310	39	0	0	250	104	73	0	0	0	0
155	584	67	63	1654	111	60	2363	162	52	3131	263	42	4096	305	39	0	0	250	103	73	0	0	0	0
160	583	66	62	1657	113	60	2362	167	52	3114	264	42	4099	300	39	0	0	250	102	73	0	0	0	0
165	581	66	62	1653	115	60	2360	173	52	3097	265	42	4103	303	39	0	0	250	101	73	0	0	0	0
170	581	65	62	1636	117	60	2359	176	52	3091	266	42	4106	306	39	0	0	250	101	73	0	0	0	0
175	586	64	62	1619	117	59	2357	179	52	3084	267	42	4110	309	39	0	0	250	100	73	0	0	0	0
180	595	65	62	1602	118	59	2356	182	52	3077	268	42	4113	311	39	0	0	250	99	73	0	0	0	0
185	601	63	62	1586	116	59	2355	185	52	3071	269	42	4115	314	39	0	0	250	98	73	0	0	0	0
190	603	63	62	1569	116	59	2353	188	52	3064	270	42	4118	316	39	0	0	250	98	73	0	0	0	0
195	605	60	61	1552	117	59	2352	191	52	3057	278	42	4121	319	39	0	0	250	97	73	0	0	0	0
200	616	61	61	1535	124	59	2350	194	52	3050	281	42	4124	319	39	0	0	250	96	73	0	0	0	0
205	637	62	61	1536	130	59	2349	197	52	3044	278	42	4126	319	39	0	0	250	95	73	0	0	0	0
210	637	62	61	1537	136	59	2347	200	51	3037	274	42	4129	319	38	0	0	250	94	73	0	0	0	0
215	638	63	61	1537	143	58	2346	199	51	3033	271	42	4135	319	38	0	0	250	93	73	0	0	0	0
220	638	64	61	1538	150	58	2345	197	51	3030	268	42	4141	318	38	0	0	250	92	73	0	0	0	0
225	638	64	61	1539	157	58	2343	196	51	3026	265	42	4146	317	38	0	0	250	92	73	0	0	0	0
230	638	65	60	1523	164	58	2342	194	51	3023	260	42	4152	317	38	0	0	250	92	73	6	0	0	0
235	639	65	60	1520	164	58	2340	193	51	3019	257	42	4198	316	38	0	0	250	92	73	0	0	0	0
240	639	66	60	1517	172	58	2339	191	51	3016	255	42	4209	315	38	0	0	250	92	73	0	0	0	0
245	637	67	60	1514	180	58	2338	190	51	3012	255	42	4220	314	38	0	0	250	92	73	0	0	0	0
250	640	67	60	1511	184	58	2336	188	51	3009	255	42	4232	314	38	0	0	250	93	73	0	0	0	0
255	641	68	60	1507	188	57	2335	187	51	3005	255	42	4243	313	38	0	0	250	93	73	0	0	0	0
260	642	69	59	1504	193	57	2333	185	51	3001	255	42	4254	312	38	0	0	250	94	73	0	0	0	0
265	642	69	59	1501	198	57	2332	184	51	2998	255	42	4265	311	38	0	0	250	94	73	0	0	0	0
270	643	70	59	1498	202	57	2331	182	51	2994	255	42	4277	311	38	0	0	250	95	73	0	0	0	0
275	644	71	59	1498	207	57	2329	168	51	2991	255	42	4288	310	38	0	0	250	95	73	0	0	0	0
280	655	71	59	1498	211	57	2328	170	50	2987	255	42	4299	309	37	0	0	250	96	73	0	0	0	0
285	666	72	59	1497	216	57	2326	200	50	2984	255	42	4299	309	37	0	0	250	96	73	0	0	0	0
290	677	73	59	1495	221	57	2325	215	50	2980	255	42	4299	309	37	0	0	250	97	73	0	0	0	0
295	689	73	58	1494	225	56	2324	222	50	2977	255	42	4299	309	37	0	0	250	98	73	0	0	0	0
300	700	74	58	1492	230	56	2322	221	50	2973	257	42	4299	309	37	0	0	250	99	73	0	0	0	0
305	711	74	58	1491	235	56	2321	225	50	2969	264	42	4318	302	37	0	0	250	101	73	0	0	0	0

Table 13, continued /gæ/ synthesis parameters

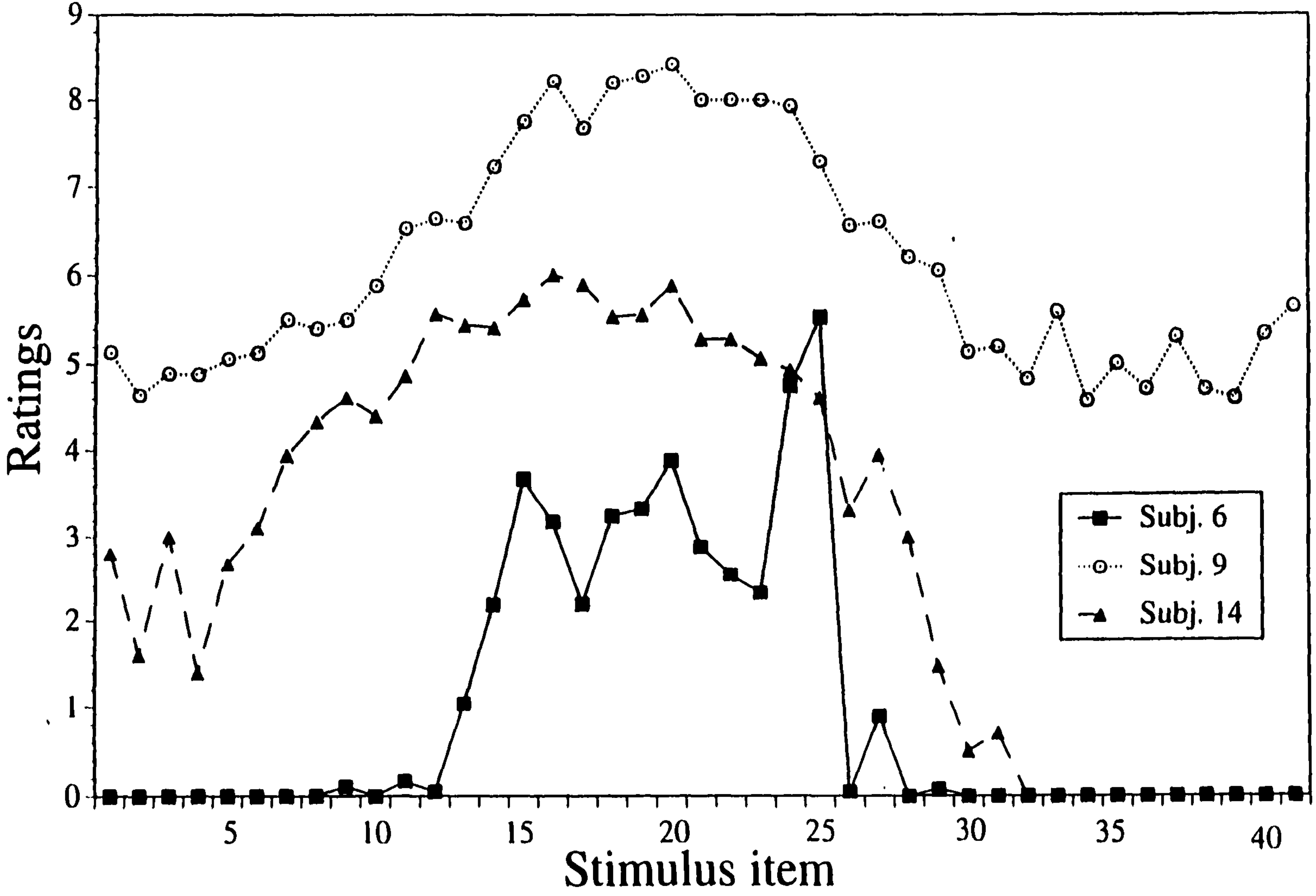
msec	F1	B1	A1	F2	B2	A2	F3	B3	A3	F4	B4	A4	F5	B5	A5	A6	AB	NZ	FO	AV	AH	AS	AN	AF
310	722	75	58	1493	239	56	2319	228	50	2966	271	42	4337	288	37	0	0	250	102	73	0	0	0	0
315	733	76	54	1495	244	53	2318	230	47	2962	278	39	4357	277	35	0	0	250	103	73	0	0	0	0
320	744	76	51	1497	248	50	2317	230	43	2959	285	35	4376	257	32	0	0	250	104	73	0	0	0	0
325	755	77	47	1499	253	47	2315	228	40	2955	291	32	4395	240	30	0	0	250	105	73	0	0	0	0
330	767	78	44	1497	258	45	2314	226	36	2952	298	29	4414	227	27	0	0	250	106	73	0	0	0	0
335	778	78	40	1490	262	42	2312	224	33	2948	305	26	4434	219	25	0	0	250	108	73	0	0	0	0
340	789	79	37	1482	267	39	2311	222	29	2945	312	22	4453	210	22	0	0	250	109	73	0	0	0	0
345	800	79	33	1475	269	36	2310	221	26	2941	319	19	4472	206	20	0	0	250	110	73	0	0	0	0

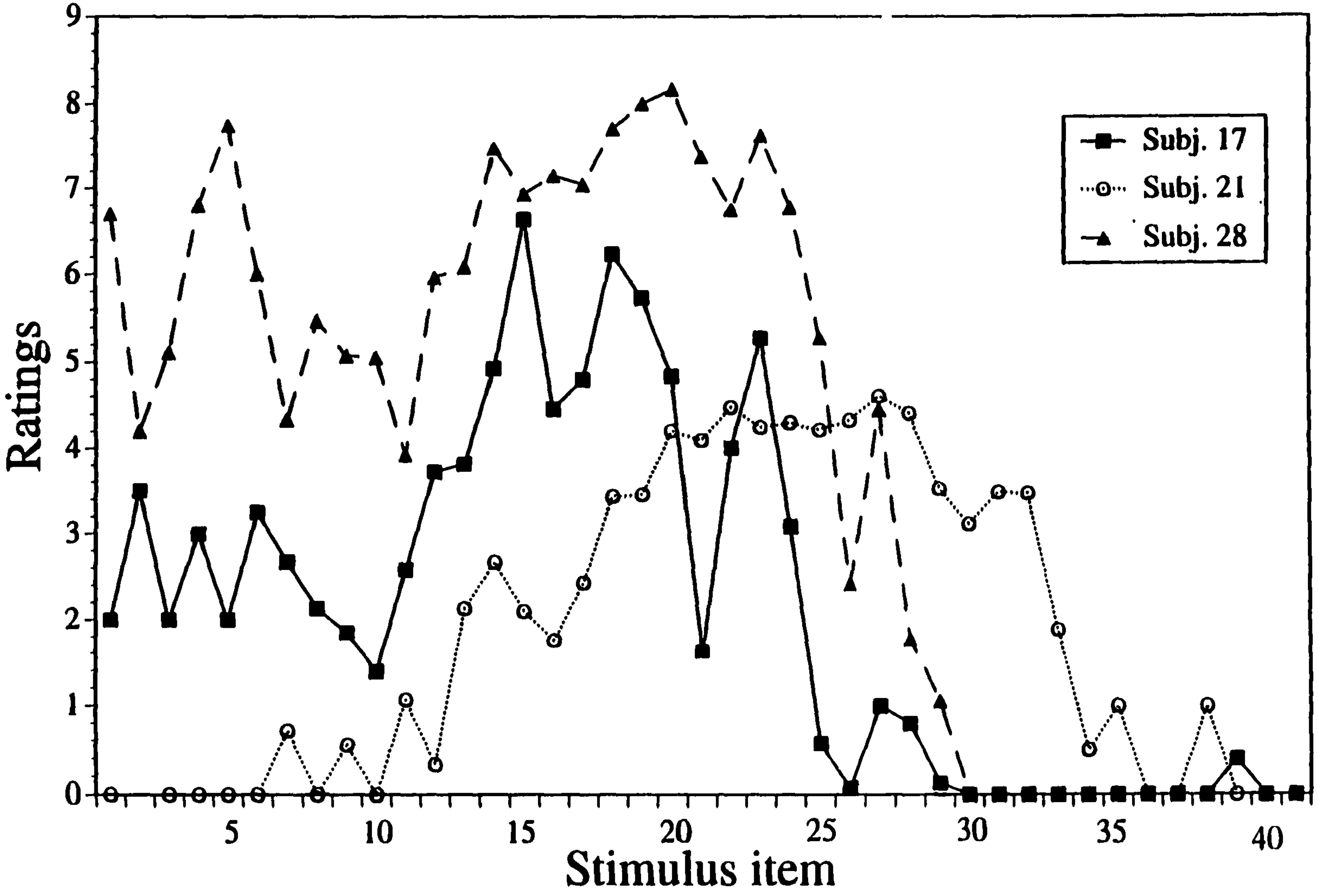
Unlike in the prior experiments, the items were presented in an adaptive testing fashion.

Results

Results were measured as in the first two experiments. For the perception task, the single item in the continuum with the highest rating was considered the listener's prototype for that dimension. Figure 15 shows the rating functions for three participants in the /b/-series session. Figure 16 and 17 likewise show rating functions for three participants in the /d/ and /g/ sessions, respectively. (Note that since this experiment used an adaptive testing method, there were fewer presentations of items that received low ratings, primarily those near the endpoints of the series).

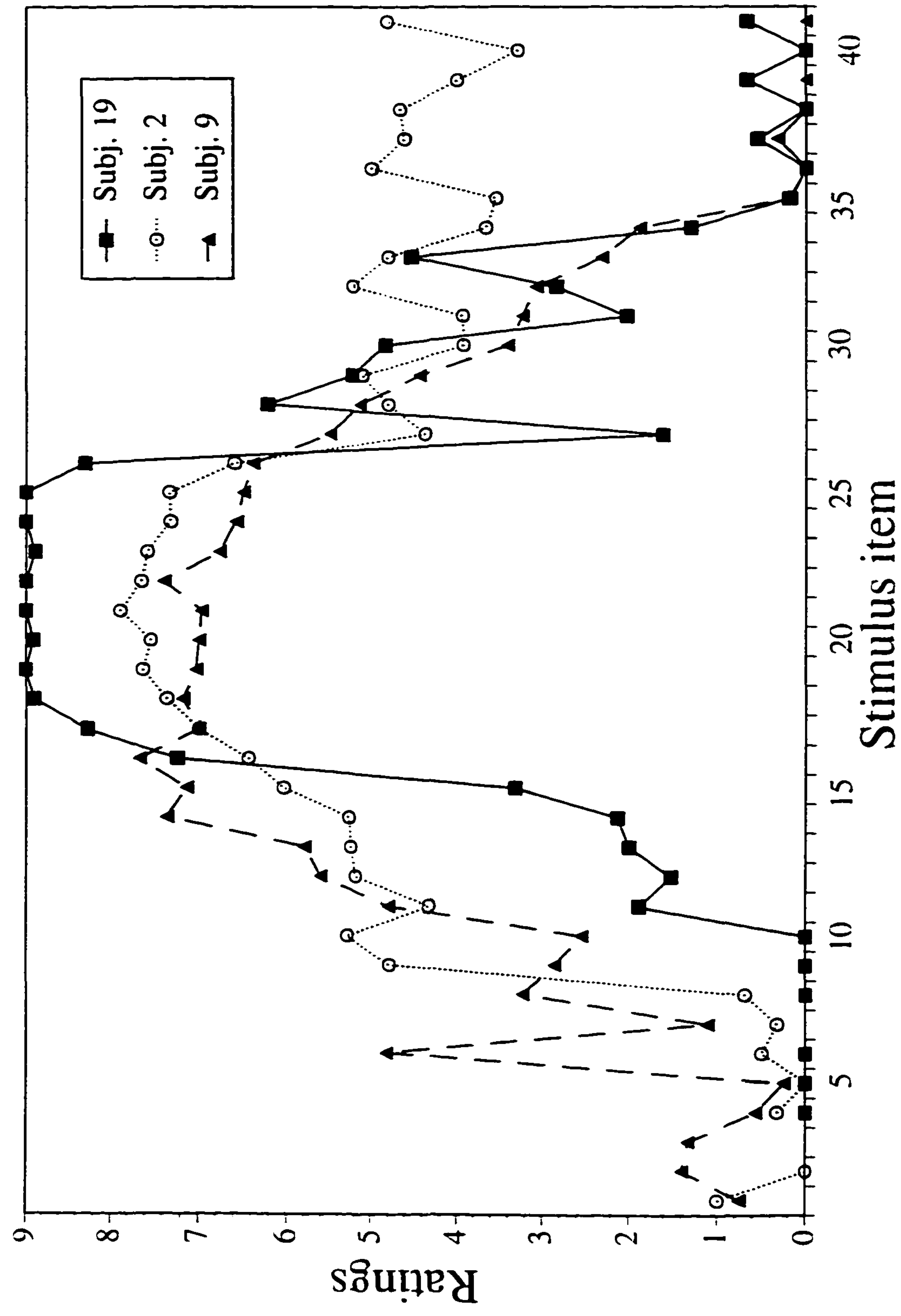
As synthetic speech sounds often are perceived differently by different individuals, subjects' data were removed from the analysis if a central member of the appropriate category could not be determined from their perceptual data. Criterion performance consisted primarily of a peak in the rating function, which received a rating of at least 4 on the 0 to 9 rating scale. Furthermore, endpoint values were required to be less than 6, and to be no more than 80% of the peak rating. Although this may have unnaturally limited the range of variability in the data, these subjects apparently did not find any of the synthetic items representative of their perceptual prototype, and inclusion of their data would have masked any


Three subjects' perceptual ratings for /bæ/



Three subjects' perceptual ratings for /dæ/

Three subjects' perceptual ratings for /gæ/



effects present. The number of subjects whose data was removed from each condition, and the reasons for this removal, will be described in more detail in the sections discussing the results with the individual phonemes.

Measurements were made of the subject's productions according to each of the metrics described above, with the exception that the spectral tilt measure was held pending examination of the other measurement results. However, unlike in the first two experiments, these measures were only taken on the tokens that phonetically matched those used in the perception task, rather than on all productions. This change was required in order to make the time requirements of the acoustic measurements more reasonable. The other recordings were saved for possible examination at a later date. All measurements were taken for each consonant separately, and in the same manner.

Measurements were made at two different points in time for each metric. For spectral peak differences, measurements were taken at the first vocal pulse, and at the first pulse occurring at least 40 ms later. For spectral moments, the initial measuring point was either the point of highest amplitude occurring in the first 10 ms of the burst + aspiration, or (if there was no burst, as was commonly the case for /b/ tokens) at the first vocal pulse. The second measurement location was identical to that for the peak differences metric. For locus equations, the first measurement was centered on the first vocal pulse. The second measurement's location was based on visual inspection of the stimulus. In order to make this inspection easier, the productions were first down-sampled to 10 kHz. If the second formant's path was shaped like an upside-down-U, the measurement was taken at the highest point in the curve. If F2 was flat, or had a linear slope, the measurement was taken at the vocal pulse midway through the course of the vowel.

For spectral moments, a spectral transformation of each stimulus was computed with a software filter bank designed to mimic Patterson's (1974) auditory filter shape. The bandwidths were similar to those for critical bands (Zwicker & Terhardt, 1980; Scharf, 1970). The frequency mean, standard deviation, skewness, and kurtosis were computed for a 15-ms temporal window centered on the peak in the spectrum (either in the burst, or the peak of the appropriate vocal pulse). These values were then averaged across the 6 productions for each speaker for each consonant. The results for each subject, as well as the measurements across subjects are given in Table 14 for /b/, /d/, and /g/.

For peak differences, the peaks were computed from a 19 ms temporal window centered on a vocal pulse. Linear predictive coding was used to find the best values for each formant. When an LPC analysis failed

Table 14 Subjects' production moments for voiced stop consonants

	b				d				g			
	mean	std. dev.	skewness	kurtosis	mean	std. dev.	skewness	kurtosis	mean	std. dev.	skewness	kurtosis
	0.264	-0.011	-0.033	-0.032	-1.621	-0.084	0.324	0.058	-1.961	0.156	0.486	-0.206
	-1.292	-0.900	0.257	0.544	-2.092	-0.627	0.328	0.252	-2.755	0.059	0.500	-0.067
	0.166	-0.375	-0.060	0.103	-2.286	-0.147	0.432	0.171	-2.828	-0.162	0.504	0.089
	0.052	-0.349	-0.093	0.037	-2.065	-0.310	0.300	0.108	-2.964	0.108	0.570	-0.036
	-0.084	-0.505	0.047	0.203	-2.998	-0.432	0.521	0.155	-2.141	-0.044	0.428	0.135
	-0.667	-0.378	0.147	0.237	-2.452	0.051	0.465	0.043	-1.633	-0.240	0.302	0.077
	-0.898	-0.481	0.170	0.326	-1.962	-0.150	0.324	0.256	-2.331	0.010	0.448	0.014
	-0.137	-0.279	-0.015	0.138	-2.212	-0.435	0.353	0.155	-2.391	0.278	0.449	0.012
	0.274	-0.419	-0.073	0.038	-2.520	-0.343	0.356	0.196	-1.893	-0.074	0.354	0.103
	-0.413	-0.208	0.139	0.108	-2.646	-0.648	0.392	0.261	-2.399	-0.129	0.410	0.063
	-1.045	-0.690	0.200	0.242	-2.180	0.216	0.313	-0.080	-2.009	-0.238	0.301	0.099
	0.094	-0.186	-0.021	-0.012	-1.369	-0.417	0.053	0.019	-1.865	0.441	0.308	-0.239
	-0.214	-0.487	0.061	0.173	-1.044	-0.278	0.063	0.087	-2.362	-0.074	0.439	0.032
	-1.430	-0.626	0.277	0.351	-2.724	-0.501	0.374	0.083	-2.546	-0.054	0.486	0.086
	-0.170	-0.215	0.040	0.067	-2.689	-0.157	0.426	0.062	-1.507	-0.398	0.217	0.096
	-0.534	-0.542	0.198	0.223	-1.885	-0.063	0.278	0.089	-1.633	0.027	0.309	-0.031
	-0.096	-0.545	-0.057	0.006	-1.731	-0.510	0.164	0.168	-2.128	0.242	0.334	-0.092
	-0.788	-0.765	0.106	0.256	-2.872	-0.468	0.472	0.241	-2.069	0.024	0.328	0.058
	0.506	-0.181	-0.145	-0.066					-2.635	-0.543	0.508	0.285
	-0.442	-0.507	0.058	0.159					-1.829	-0.352	0.225	0.052
	-0.118	-0.573	-0.109	0.142					-2.194	-0.088	0.349	-0.037
	-0.078	-0.396	0.051	0.214					-2.204	-0.022	0.396	0.119
	0.659	-0.296	-0.234	-0.166								
	0.269	-0.441	-0.060	0.007								
AVG.	-0.255	-0.431	0.035	0.137	-2.186	-0.295	9.432	0.129	-2.194	-0.049	0.393	0.028

to find a peak, a narrow-band spectrum (using a 24 ms window)[17] was used instead. In the few cases when neither method was capable of finding a missing peak, the average value of that formant for the other 5 tokens of the same syllable was inserted.

The peak values were then converted into their Bark scale equivalents (Zwicker & Terhardt, 1980). The Bark scale was used because it gives a more accurate representation of the processing abilities of the human auditory system. The difference scores were calculated between the first peak and the fundamental frequency (p1-f0), the first and second spectral peaks (p2-p1), and between the second and third (p3-p2), the third and fourth (p4-p3), and the second and fourth (p4-p2). These values were then averaged across the 6 productions for each speaker for each consonant. The average values are given in Table 15 for each speaker and averaged across speakers at the bottom of the table.

Locus equations (by definition) are based on change in F2 over a wide variety of contexts. Because this experiment involves measuring transitions in only one vowel environment for each consonant, it is not, strictly speaking, appropriate to determine slopes and y-intercepts from these values. Furthermore, because the perceptual task results in only one

---

[17] There is a tradeoff between temporal resolution and frequency resolution. Thus, in order to get better frequency resolution, it is necessary to use a larger temporal window (and thus lose some degree of temporal precision).

Table 15 Subjects' production peak differences for voiced stop consonants

	b					d					g				
	Δp1-f0	Δp2-p1	Δp3-p2	Δp4-p3	Δp4-p2	Δp1-f0	Δp2-p1	Δp3-p2	Δp4-p3	Δp4-p2	Δp1-f0	Δp2-p1	Δp3-p2	Δp4-p3	Δp4-p2
	1.091	-0.007	-0.925	0.047	-0.877	1.109	-0.883	-0.404	-0.163	-0.567	1.217	-1.456	0.306	-0.240	0.066
	1.786	-1.458	-0.219	0.021	-0.198	1.692	-1.386	-0.126	-0.324	-0.450	1.594	-2.002	0.344	0.251	0.595
	0.891	0.106	-0.452	-0.184	-0.637	1.519	-1.333	-0.197	-0.254	-0.451	1.853	-2.341	-0.078	0.420	0.343
	0.917	-0.275	-0.160	-0.259	-0.419	1.553	-1.428	-0.112	-0.011	-0.123	1.213	-1.574	0.366	-0.017	0.349
	1.737	-1.252	-0.074	-0.043	-0.117	1.822	-1.671	-0.219	-0.063	-0.282	2.079	-2.306	-0.199	0.385	0.186
	0.106	0.527	-0.432	0.058	-0.374	1.238	-1.089	-0.118	-0.162	-0.280	1.488	-1.836	0.686	-0.304	0.382
	1.239	-0.935	-0.214	-0.031	0.245	1.125	-0.970	-0.390	0.061	-0.329	1.578	-2.091	0.097	0.440	0.537
	0.827	-0.291	-0.302	-0.089	-0.391	1.285	-0.953	-0.272	-0.373	-0.644	1.275	-1.537	-0.035	0.259	0.224
	1.175	-0.428	-0.444	-0.052	-0.496	1.857	-1.869	-0.022	0.014	-0.008	2.365	-3.252	0.615	0.151	0.766
	1.087	-0.686	0.360	-0.404	-0.044	1.398	-1.458	-0.281	0.392	0.111	1.839	-2.462	0.572	0.099	0.671
	1.021	-0.593	-0.026	-0.140	-0.166	0.925	-0.610	-0.382	-0.378	-0.760	1.393	-1.540	-0.146	0.542	0.397
	1.374	-0.064	-0.756	-0.379	-1.128	0.996	-0.643	-0.514	-0.196	-0.710	1.148	-1.340	0.080	-0.034	0.046
	1.763	-1.513	0.330	-0.609	-0.279	1.321	-0.882	-0.471	-0.326	-0.797	1.072	-1.157	-0.506	0.348	-0.158
	1.529	-0.959	-0.485	-0.553	-1.038	2.172	-2.169	-0.211	0.008	-0.204	1.618	-2.340	0.391	0.161	0.552
	0.564	0.456	-0.547	-0.205	-0.752	1.027	-0.734	-0.147	-0.203	-0.349	1.826	-2.320	0.221	0.210	0.479
	1.280	-0.577	-0.171	-0.258	-0.429	1.436	-1.297	-0.323	-0.063	-0.386	1.140	-1.547	0.501	-0.148	0.353
	1.073	-0.702	0.070	-0.370	-0.299	1.110	-0.551	-0.504	-0.025	-0.529	1.432	-1.773	-0.266	0.411	0.145
	0.879	-0.198	-0.308	-0.030	-0.339	1.816	-1.632	-0.145	0.030	-0.115	1.640	-2.447	0.280	-0.089	0.191
	1.033	-0.242	-0.334	-0.120	-0.454						1.872	-2.484	0.593	0.433	1.026
	1.474	-1.168	0.155	-0.122	0.033						1.980	-2.443	-0.007	0.536	0.529
	1.892	-1.159	-0.284	-0.264	-0.547						1.254	-1.614	0.170	0.124	0.294
	0.942	-0.137	0.501	-0.666	-0.165						1.437	-1.990	0.090	0.618	0.708
	1.090	-0.106	-0.632	-0.126	-0.759										
	1.140	-0.579	-0.242	0.116	-0.126										
AVG.	0.163	-0.510	-0.233	-0.194	-0.407	1.411	-1.198	-0.269	-0.113	-0.382	1.560	-1.993	0.185	0.207	0.395

value for each subject, it is impossible to find slopes and y-intercepts perceptually. Thus, rather than examine the locus equations *per se*, the current experiment examined the change in the second formant (ΔF2) for each subject instead. As this is the primary information upon which locus equations are calculated, this switch should still allow the investigation of the correlation in locus equations across perception and production. That is, if the changes in F2 are not highly correlated across the two modalities, the locus equations would likewise not be highly correlated. F2 measurements were taken in the same manner as for the peak differences, except the values were not then transformed into their Bark equivalents. The value at consonant onset was subtracted from the value found midway through the vowel, and these difference scores were then averaged across the 6 productions for each talker. These average values are given in Table 16 for each talker, and, at the bottom, across talkers for /b/, /d/, and /g/.

For the locus equations, a correlation was taken between the change in F2 in each participant's spectral prototype and the average change in F2 in their productions. Unfortunately, there is no well-accepted statistical test for calculating the overall correlations between sets of values, making the testing more difficult for the spectral moments and peak differences values. To get around this difficulty, two sets of correlations were taken. First, individual correlations were taken for each submeasure. Thus, for

Table 16

Average changes in F2 for individual subjects

	/b/	/d/	/g/
	374	-205	-383
	95	-26	-254
	165	-39	-373
	166	45	-425
	-20	-78	-593
	261	-24	-559
	-69	8	-269
	98	-63	-267
	343	-309	-692
	80	-72	-366
	65	-9	-328
	382	-99	-378
	-33	-71	-110
	168	-58	-455
	210	-68	-565
	326	12	-395
	77	307	-343
	-102	-254	-143
	198		-275
	99		-362
	219		-366
	104		-613
	308		
	-67		
	----	----	-----
AVERAGE	144	-56	-387

the moments data, the frequency mean for production was correlated with the mean for perception. The standard deviations were then correlated with one another, independently from the means, as were the values for the skewness and kurtosis. For the peak differences data, correlations for each of the 5 peak differences were likewise calculated.

Although these four (or five) correlational values give some sense of the individual relationships between members of a set, they do not give any overall correlations between sets as a whole. As peak differences and moments each have been proposed as a set of values, there is no reason to believe that the individual members would of necessity correlate with one another. That is, if each component is a dimension in multi-dimensional space, the overall location of a value in space would depend on the values for all four (or five) measures, but need not correlate highly with any single measure. Thus, in order to get some notion of overall correlational values, a canonical correlation was performed. This test correlates a set of independent variables (IVs) with a set of dependent variables (DVs). However, it does so by searching for the linear combinations of IVs that best predicts a linear combinations of DVs. This method of searching gives a multiplicity of separate canonical correlations, rather than a single, overall measure of the strength of the relationship. Interpreting the relationship between the IVs and DVs can be difficult, as it depends on the factor loadings or weights for each item (that is, on how much each IV and DV contributes to the overall combination) (see Cohen & Cohen, 1983). Lastly, in order to achieve a likelihood of statistical significance, canonical correlation requires a minimum of 10 subjects per IV. Thus, for the moments data, a minimum of 40 subjects would be needed, and for the peak differences data, a minimum of 50 participants would be required. Given the difficulty of acquiring measurements from this many subjects, the results from a canonical correlation are unlikely to reach significance, even when the relationship is quite strong. However, as there is currently no well-accepted alternative, I decided to perform a canonical correlation, and examine the values for the first canonical correlate. Although several correlates might actually be present, the first (or "best" correlate) will provide some sense of the overall correlations between sets. It is important to bear in mind, however, that high correlations might not reach statistical significance, given the low *n*. Thus, results from this analysis are best considered to be exploratory, rather than conclusive, and to give suggestions of areas in which further research might be important.

Correlations between the /bæ/ production and perception measures, /dæ/ production and perception measures, and /gæ/ production and perception measures were calculated for each of the three metrics described above. The results from each of these phonemes are discussed separately.

Perception and production of /b/

A number of subjects had to be dropped from the analysis. Two subjects started recording too soon during the production task, and consequently the onsets of their productions were cut off, preventing their measurement. Data from 8 subjects were dropped for failure to reach criterion responding in the perceptual task. (Of these, 5 had to be dropped from all three portions of the experiment. It is possible these participants may have misunderstood the experiment, or, perhaps more likely, may have simply felt that all of the synthetic stimuli were poor-sounding, and thus given them all relatively low ratings. As any subject whose average peak ratings was not higher than a 4 was dropped from analysis, rating all items as relatively poor-sounding would have resulted in a failure to reach criterion.) This left a total of 23 subjects in this part of the experiment.

The change in F2 frequency had a marginally significant correlation of .378 between perception and production ($z=1.825$, $p <.07$). Although non-significant, this result is high enough to be suggestive, if a similar result is found with the /d/ and /g/ portions of the experiment.

The individual correlations from the moments data and peak differences data were less encouraging. For the moments, there were no significant or marginal correlations: for the change in mean, r=0.143 (z=0.659, p >.50); for the change in standard deviation, r=0.320 (z=1.518, p >.12); for the change in skewness, r=0.002 (z=.008, p >.99); for the change in kurtosis, r=0.269 (z=1.265, p >.20). For the peak differences, there were no significant correlations, and only one marginal correlation (but in the opposite direction): for the change in p1-f0, r=-0.378 (z=-1.823, p <.07); for the change in p2-p1, r=0.023 (z=0.105, p >.91); for the change in p3-p2, r=0.332 (z=1.581, p >.11); for the change in p4-p3, r=0.142 (z=.657, p >.51); and for the change in p4-p2, r=0.181 (z=0.839, p >.40). Even leaving aside the issue of significance, only 4 of these correlations would account for at least 10% of the variability: the change in F2 over time, the change in standard deviation, the change in p1-f0, and the change in p3-p2.

The canonical correlation results, however, are much stronger. For peak differences, the first canonical variable was significant (Chi-square = 38.88, p <.007, indicating that at least one variable is necessary to express the dependency between sets). The correlation was 0.80, explaining 64% of the variability. For the moments, the first variable was marginally significant (Chi-square = 25.84, p <.06), with a correlation of 0.83 (explaining 68% of the variance). This suggests that while the individual peak difference and moments scores may not correlate well between perception and production, the pattern represented by the set of values on each metric does seem to correlate across individuals. Interestingly, the correlations are quite similar for the peak difference and moment data. If this holds for the /d/ and /g/ productions as well, it might suggest that both sets of variables are related to the cues people actually use, but that neither set is related any more closely than the other. That is, neither set is entirely accurate, although both sets correlate with the cues people use.

Perception and production of /d/

As with the /b/ data, a number of subjects had to be dropped from the analysis. Data from 16 subjects were dropped for failure to reach criterion responding in the perceptual task (including the 5 already mentioned whose data were dropped from all three portions), leaving a total of 17 subjects. A much larger proportion of subjects apparently had difficulty with the synthetic /d/ stimuli than with the /b/ stimuli. This is worrisome, and calls into question the generalizability of results from the remainder of the subjects.

Leaving aside for the moment the issue of generalizability, the results from the correlations were no more impressive than those from the /b/ data. The change in F2 frequency had a nonsignificant correlation of .241 between perception and production ($z=0.954$, $p > .34$), similar to the null result found by Ainsworth and Paliwal (1984) for F2 and F3 loci. For the moments, there were no significant or marginal correlations: for the change in mean, r=0.059 (z=0.231, p >.81); for the change in standard deviation, r=-0.243 (z=-0.962, p >.33); for the change in skewness, r=0.008 (z=.031, p >.97); for the change in kurtosis, r=-0.220 (z=-0.867, p >.38). For the peak differences, there were likewise no significant or marginal correlations: for the change in p1-f0, r=0.323 (z=1.296, p >.19); for the change in p2-p1, r=0.323 (z=1.296, p >.19); for the change in p3-p2, r=-0.310 (z=-1.243, p >.21); for the change in p4-p3, r=-0.109 (z=-0.424, p >.67); and for the change in p4-p2, r=0.069 (z=0.269, p >.78). Again setting aside the issue of significance, only 2 of these correlations would account for at least 10% of the variability: the change in p1-f0 (which was similarly high for the /b/ items, but in the opposite direction), and the change in p2-p1.

The canonical correlation, results, however, are much stronger. Although no variables were significant (not surprising given the small *n*), the correlations for both the peak differences and the moments were 0.69 (explaining 48% of the variance). As both sets of cues provide equivalent correlations, it suggests that the cues listeners actually use are related to both of these aggregate sets equivalently.

Perception and production of /g/

As with the /b/ data, several subjects had to be dropped from the analysis. One subject started recording too soon during the production task, and consequently the onsets of her productions were cut off, preventing their measurement. Data from 10 subjects were dropped for failure to reach criterion responding in the perceptual task (including the data from the five participants who failed to reach criterion in any portion of the experiment). This number is more in line with the data from /b/ than /d/ results, but still constitutes a fair number of subjects. This left data from a total of 22 subjects in this portion of the experiment.

The results from the correlations were similar to those from the /b/ and /d/ data. The change in F2 frequency had a nonsignificant correlation of -.027 between perception and production ($z=-0.119$, $p > .90$). Thus, for the three consonants, two showed non-significant locus correlations, and one showed a marginal correlation.

For the moments, there were no significant or marginal correlations: for the change in mean, $r=-0.115$ ($z=-0.506$, $p > .61$); for the change in standard deviation, $r=0.024$ ($z=0.103$, $p > .91$); for the change in skewness, $r=0.173$ ($z=.764$, $p > .44$); for the change in kurtosis, $r=0.029$ ($z=0.126$, $p > .89$). For the peak differences, there was one significant correlation: for the change in p1-f0, $r=0.448$ ($z=2.103$, $p < .04$). This is certainly suggestive. However, given the large number of correlational tests performed, it is likely that at least one correlation would have been significant by chance alone. With a Bonferroni adjustment for the 10 correlations, an alpha level of .005 would be required for significance, which the correlation on changes in p1-f0 does not reach.

No other correlations reached significance: for the change in p2-p1, r=0.165 (z=0.725, p >.46); for the change in p3-p2, r=-0.048 (z=-.208, p >.83); for the change in p4-p3, r=0.069 (z=0.300, p >.76); and for the change in p4-p2, r=0.227 (z=1.006, p >.31). Again setting aside the issue of significance, only 1 of these correlations would account for at least 10% of the variability: the change in p1-f0 (which was similarly high for the /d/ and /b/ items, although in the opposite direction for the /b/).

The canonical correlation results are fairly strong. As with the /d/ productions, no variables were significant given the small *n*, but the correlation for the peak differences was 0.78 (explaining 61% of the variance), and for the moments was 0.74 (explaining 55% of the variance). Again, the differences between sets of measures was very slight, but (as with the /b/ productions), the peak differences correlation was slightly higher. This difference, however, is likely too small to be of theoretical importance. Rather, it appears that listeners use neither the peak differences, nor the moments, to distinguish stop consonants, but rather use some other cue or cues that contains some of the same information. Alternatively, listeners could be making use of redundancies in the signal and using both sets of information (see Richardson, 1992).

Comparisons across phonemes

A separate issue from that of perception-production correlations is whether these sets of values could potentially be used for discriminating consonants. One way to investigate this is to determine whether there are significant differences between the values for each of the three consonants. For this analysis, rather than include differing numbers of subjects in the three conditions, only data from those 15 subjects who reached criterion in all three conditions were used. Previous research (Richardson, 1992) has suggested that means and standard deviations are the most critical of the four moments data for distinguishing on place of articulation, and that p1-f0 and p3-p2 are the most critical of the five peak-difference values. Only this subset was tested here. An overall ANOVA compared the differences between /b/, /d/, and /g/ for the 5 measures of change in p1-f0, p3-p2, mean, standard deviation, and F2 (locus). This suggested that there was an overall difference in the phonemes ($F(2,28)=113.061$, $p <.0001$). There was also an overall effect of cue (caused presumably by the fact that the values for F2 differences were approximately two orders of magnitude larger than the values for the changes in Bark values for mean and peak differences; $F(4,56)=11.978$, $p < .0001$). There was also a significant interaction ($F(8,112)=112.745$, $p < .0001$). Follow-up t-tests were used to determine where significant differences lie. The requirement that the ANOVA be significant should protect against an inflated alpha level, even with a large number of statistical tests. However, to be conservative, the alpha level was lowered to .0033, to adjust for this number (15) of statistical tests, according to Bonferroni's approach. The t-tests suggested that the mean value for /b/ productions was different from that of /d/ and /g/ productions, but the latter two did not differ (b vs. d: $t\ (14)=13.881$, $p < .0001$; b vs. g: $t\ (14)=12.323$, $p < .0001$; d vs. g: $t\ (14)=-0.684$, $p > .50$). The standard deviations were different for /g/ than for /b/ and /d/ productions, which did not differ (b vs. d: $t\ (14)=-1.943$, $p > .07$; b vs. g: $t\ (14)=-6.863$, $p < .0001$; d vs. g: $t\ (14)=-9.179$, $p < .0001$). Combined, then, these two moment values would serve to differentiate all three places of articulation (/b/ tends to have a much larger mean than the other two, /g/ has a much larger standard deviation, and /d/ has relatively small values on both measures.) The degree of change in F2 differentiated all three consonants (b vs. d: $t\ (14)=5.517$, $p < .0001$; b vs. g: $t\ (14)=13.301$, $p < .0001$; d vs. g: $t\ (14)=11.528$, $p < .0001$). The change in p1-f0 was different for /b/ than for /g/ productions ($t\ (14)=-5.951$, $p < .0001$) but only marginally different for /b/ vs. /d/ ($t\ (14)=-3.392$, $p > .004$); there was no difference between /d/ and /g/ productions (*t* (14)=-0.773, *p* >.45). The change in p3-p2 was different for /g/ than for either of the two other places of articulation (b vs. g: *t* (14)=-4.506, *p* =.0005; d vs. g: *t* (14)=-5.856, *p* <.0001), but the /b/ and /d/ did not differ from one another (*t* (14)=-0.990, *p* >.33). These results suggest that the F2 and moments data could be used to differentiate the three places of articulation, even though it is not entirely clear from the perception/production correlations that subjects actually did so. The peak differences data might also be used, although it might be more difficult to differentiate /b/ from /d/ productions using just the P1-F0 and P3-P2 dimensions of this metric.

Conclusions

It appears that any of the proposed sets of cues could be used by listeners to distinguish the different places of articulation. However, if we assume that perception-production correlations can be used to evaluate the usefulness of different cues, none of these sets seem to adequately portray what listeners actually do.

It may simply be that perception-production links cannot be used to evaluate perceptual cues in this manner. However, the high canonical correlations for both moments and peak differences seems to suggest that this method may be able to pick out the relative usefulness of a cue. If so, it suggests that both of these sets of cues are used equivalently, to the extent that they are used at all. Given the variability in the prior literature, this ambiguous result may not be that surprising. Perhaps the best conclusion is that listeners are using a set of cues that has not yet been formally suggested in the literature, but which seems to include some of the same information included in the moments and peak differences descriptions. That is, the real cue listeners use is neither set, but rather something related to both sets.

The poor correlation for the F2 locus value is less heartening. Unfortunately, there is no way to evaluate locus equations directly, as these require multiple values (something impossible to determine from a single perceptual prototype). It is unclear whether a higher correlation would have been found if there was some way of evaluating locus equations, rather than individual locus values. Given this uncertainty, perhaps the only conclusion that can be made is that there is no apparent evidence for the use of locus values as a cue based on the correlation between perception and production.

CHAPTER 6

Concluding Remarks

In the first experiment, I examined the link between perception and production in a series varying in voice onset time (VOT). The data suggested that people who produced the token /pɑ/ with a longer VOT also had perceptual prototypes of /pɑ/ with a longer VOT. That is, there was a correlation between the individual prototypes in perception and the average VOTs in production. Furthermore, the production of /ba/ also correlated with the VOT of the /pɑ/ prototype, and explained additional variance beyond that of the /pɑ/ production. This suggests that the VOT of voiced tokens in production is at least partly independent from the VOT of voiceless tokens (that is, that individuals who produce long VOTs in their voiceless items do not necessarily produce relatively long VOTs in their voiced items), and that this separate production factor nonetheless is correlated with perception. In addition, there was some evidence to support Johnson *et al.*'s claim that perceptual representations are hyperarticulated, since individual's preferred VOTs that were more extreme than their own productions.

The results from this first experiment suggest that there is a link between perception and production. However, the second experiment results did not support this. This second experiment examined series ranging from /s/ to /ʃ/, and varying in either frication centroid or in the formant values at frication offset. Frication is viewed as the primary cue distinguishing /s/ from /ʃ/, and was predicted to result in a larger perception-production correlation than was the formant cue (which is viewed as a secondary cue, at best). However, there were no significant correlations between perception and production on either the frication or the formant measures.

In the third experiment, correlations between perception and production were examined on the basis of three different cues in three different series. Series based on /b/, /d/, and /g/ were presented for goodness ratings, and perceptual prototypes were found for each series. Both these prototypes and subjects' productions of /bæ/, /dæ/, and /gæ/ were analyzed for their F2 loci, peak differences, and spectral moments. There was no consistent correlation between the F2 loci in perception and production, and nor were there significant correlations between perception and production of any of the individual measures making up spectral moments or peak differences. However, looking at the sets of different measures making up moments and peak differences, there were some trends towards perception-production relationships. Canonical correlations (examining these sets of measures) found fairly high values of r for both the moments and the peak differences measures. Unfortunately, the large number of subjects required by canonical correlations made it impossible to examine the statistical significance of these findings. Therefore, these results must be viewed as tentative at this point. Furthermore, the correlations were nearly identical for the spectral moments and peak differences data, providing no hint as to which set of cues might be more strongly related to the cues actually used in perception. Perhaps both sets of cues are used in a highly-redundant system. Or, perhaps neither set accurately describes the cues listeners actually use on-line, and both sets are equivalently related to the "real" cues. It is impossible to distinguish between these possibilities at this point.

In general, then, the results from these experiments are less clear than desired. However, a few key points do appear. The basic question behind these experiments was whether individual differences in perception might be correlated with individual differences in production. That is, whether perception and production are linked at the level of the individual talker/listener. The results from Experiment 1 suggest that this is the case. Individuals whose perceptual prototypes for the sound /p/ have longer voice onset times also had longer VOTs when producing this phoneme. Although the results from later experiments failed to uphold this basic finding, it is worth noting that the cue used in Experiment 1 (VOT) is likely the most accepted cue proposed in the literature. There is more evidence supporting the use of VOT in perception than for any other cue. On the other hand, the cues described in Experiment 3, which led to fairly ambiguous results, are perhaps the proposed cues most in contention. There have likewise been alternative proposals for measuring the frication and formant cues used in Experiment 2. This may explain why only Experiment 1 has led to significantly positive results. Perhaps finding correlations between perception and production depends critically on examining a cue that listeners actually use during their on-line recognition of phonemes. If so, it would suggest that frequency centroids for fricatives and the spectral moments and frequency differences between spectral peaks for voiced stops are all inaccurate descriptions of listeners' perceptual cues.

On the other hand, it may also be the case that perception-production correlations are relatively slight, such that any large degree of variability in measurement makes them difficult to find. Or perhaps they are only present for certain types of phonetic distinctions. The latter would bring into question the whole notion of linkages between the input and output modalities, as any overall connection between them should be independent of phonetic identity. Unfortunately, the results from the current set of experiments make it difficult to decide between these alternative explanations. There is no evidence from the current sets of experiments to suggest that perception-productions links can be found outside of VOT continua, although there are alternative explanations for the failure to find significant effects in Experiments 2 and 3.

Regardless, it appears unlikely that examining perception-production correlations will be of use in helping to distinguish between alternative sets of proposed cues. Many proposed “cues” are actually sets of cues, and the large numbers of subjects required by canonical correlations make examination of these metrics difficult. For these cues, evidence from perception-production correlations is unlikely to be worth the effort it would entail.

The present results have a number of theoretical implications. The mixed findings, however, make interpretation difficult. As has already been discussed, it is unclear whether the lack of effects in the second and third experiment were caused by an inappropriate measure or by a true absence of an effect. Coarticulation can make the choice of an acoustic measure difficult, and it is possible that VOT is the only appropriate cue used in this set of experiments. This makes it impossible to entirely rule out any potential causes of a perception/production link. Such a link could theoretically be mediated by several sources. The most extreme view is that perception and production both involve the same mental representations. This is the view proposed by motor theory, for example (Liberman et al., 1962; Liberman et al., 1967; Liberman & Mattingly, 1985). However, if this were the case, correlations between perception and production should always be present, assuming a proper procedure and appropriate measure are used. The pattern of results in the current set of studies, as well as in the prior literature, suggest that finding these correlations is not a trivial matter. The correlations can be found in some instances, but they do not appear to be entirely consistent, nor readily apparent in all cases. However, as stated above, it is possible that this variability in results is because of a failure to find an appropriate acoustic measure, rather than because of a small, variable correlation. Thus, there is still some room for contention with regards to this theory.

An additional argument against the same-representation idea comes from the work of Johnson *et al.* (1993). They found that representations seem to be more extreme in perception than in production. This finding has been replicated by Freida (1997) for vowels, and has also been supported by results from Experiment 1. If representations are more extreme perceptually than in production, it would necessitate that these representations be separate, arguing against motor theory. However, it is possible that participants in these experiments rated items not for their typicality, but for their distinctiveness, especially since the items were not in a normal, fluent speech context. That is, individuals may have interpreted the instructions as meaning that they should judge items on the basis of how easily they could be distinguished from other phonemes, rather than judging them as to their normalcy. Thus, the hyperarticulation effect could be caused by task factors, rather than by representational differences. This makes it impossible to dismiss the view that perception and production involve the same mental representations, although the current results do not provide much support for such a theory.

A second possibility is that while the representations are not identical, they are directly connected in some manner. This would suggest that changes in one representation should cause similar changes in the other, but that the two representations need not be identical. Although this would allow for the hyperarticulation results of Johnson *et al.*, and of Experiment 1, it would still suggest that correlations between these representations should be relatively straight-forward to find, assuming a correct task and perceptual measure. Again, the mixed current results are not able to rule out this theory, since it is possible that an inappropriate measure was used in Experiments 2 and 3. However, the results do not provide much support for such a view, either.

Another possibility is that the representations are distinct, but that the perceptual prototype is based on exemplars, weighted according to their frequency of occurrence. That is, individuals' idealized perceptual expectations are based on all of the instances of a sound that they have heard up to that point in time. Since people are likely to have heard their own productions more than that of any other single individual, their productions are likely to have an especially important role in their perceptual prototypes. A closely related proposal is that these prototypes are based on all of the instances of a sound the individual experienced until some critical point in their childhood, but is less influenced by examples heard thereafter. Either of these proposals would fit well with theories of speech perception such as Fowler's gestural-based theory (1986) and Nearey's double weak theory (1992)

According to either of these similar points of view, the link between perception and production is indirect. A person's own productions would have a prominent role in the development of that individual's perceptual prototypes, but would not be the only critical factor. Thus, perceptual expectations should be a skewed towards one's own productions, but other individuals the listener has heard frequently would have a similarly high contribution to his or her perceptual prototypes. This might suggest that listeners' perception would be correlated not only with their own production, but also with the productions of family members and close friends. Although this prediction is testable in theory, it may be less so in practice. Since children model their productions on the basis of what they hear around them, their productions are likely to be highly correlated with the productions of parents and caretakers.[18] This may make it difficult to find a correlation between an individual's perception and her primary caretaker's production over and above the correlation between the individual's perception and her own production, at least for normal speakers.

This may be less of a problem for disordered speakers, however. For example, children with cleft palate have great difficulties producing certain classes of phonemes. One such difficulty is that they frequently produce voiceless stops with far longer VOTs (voice onset times) than are produced by normal speakers. The exaggerated VOTs these children produce, even after surgical intervention would allow them to produce sounds normally, makes it far more likely that their productions do not correlate very highly with their parents' productions. In addition, there is a known etiological cause for these children's articulation difficulties, unlike the misarticulating children discussed in Chapter 1. This allows us to be fairly certain that the disordered production is not caused by any underlying perceptual disorder. Plus, VOT seems to be the one perceptual cue for which perception-production links can be found with some success in normal speakers. This would provide the opportunity, then, to examine

---

[18] There is some anecdotal evidence in favor of such a view. Some school teachers have reported finding children of hearing-impaired parents who demonstrate no hearing loss themselves, but who articulate speech in a manner akin to their hearing-impaired parents (Mara Boettcher, 1996, personal communication).

the relative influence of individuals' own productions and of their parents' productions on their perceptual prototypes. If these children show no correlation between their own production and perception, it would suggest an ability to discount their own aberrant productions, and would provide further evidence against the notion of a combined perception/production representation. If the children show correlations between their perception and their production, but no additional correlation between their perception and their parents' productions, it might suggest that perceptual representations are determined solely by the single voice most often heard, and are not influenced by other frequently-heard voices. This would also provide some support for a more direct connection between production and perception. If, on the other hand, both the children's and their parents' productions correlate with their perception, it would provide strong evidence in favor of an exemplar-based (or prototype) representation in which the perceptual representations are determined by experience, with the voices heard most frequently having the largest influence.

In Chapter 1, I suggested that a correlation between perception and production would be difficult to reconcile with connectionist theories such as TRACE. This was because the presence of direct links between perceptual and productive representations would change the nature of the model as a whole. However, the ambiguous results from the present set of experiments seem most supportive of a model with only indirect connections between the modalities, as in the exemplar model discussed above. This type of "link", for lack of a better word, would not necessarily pose difficulties for TRACE. Thus, the present results do not seem to rule out this type of model.

In fact, even though the results from the first experiment seemed to support the idea of motor theory at the expense of numerous other proposals, the results from the set of experiments as a whole may actually have the opposite implication. That is, these results seem to suggest that any connections across perception and production are indirect. This finding can be accommodated by all models except for motor theory.

In conclusion, there is some evidence for perception-production correlations, at least for some contrasts. However, these correlations are somewhat difficult to find, which argues against the notion that the representations are actually identical in the two modalities. In fact, these results seem to best fit a model which has no direct link between perception and production at all. Correlations between the representations used in perception and production can be explained by the fact that the voice that one has the most experience with and which one hears the most often is one's own. This familiarity can cause a skewing of perceptual expectations towards one's own voice, while still maintaining a modular structure in which perception and production are entirely separate structures.

References

Ades, A. (1974). Bilateral component in speech perception? Journal of the Acoustical Society of America, 56(2), 610-616.

Ades, A. (1977). Source assignment and feature extraction in speech. Journal of Experimental Psychology: Human Perception and Performance, 3(4), 673-685.

Ainsworth, W. A. (1977). Mechanisms of selective feature adaptation. Perception & Psychophysics, 21, 365-370.

Ainsworth, W. A., & Paliwal, K. K. (1984). Correlation between the production and perception of the English glides /w, r, l, j/. Journal of Phonetics, 12, 237-243.

Aungst, L. F., & Frick, J. V. (1964). Auditory discrimination ability and consistency of articulation of /r/. Journal of Speech and Hearing Disorders, 29, 76-85.

Bailey, P. J., & Haggard, M. P. (1973). Perception and production: Some correlations on voicing of an initial stop. Language and Speech, 189-195.

Bailey, P. J., & Haggard, M. P. (1980). Perception-production relations in the voicing contrast for initial stops in 3-year-olds. Phonetica, 37, 377-396.

Behrens, S. J., & Blumstein, S. E. (1988). Acoustic characteristics of English voiceless fricatives: a descriptive analysis. Journal of Phonetics, 16, 295-298.

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Sawusch, J. R. (1979). Some relationships between speech production and perception. Phonetica, 36, 373-383.

Blumstein, S. E., Isaacs, E., & Mertus, J. (1982). The role of the gross spectra shape as a perceptual cue to place of articulation in initial stop consonants. Journal of the Acoustical Society of America, 72, 43-50.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. Journal of the Acoustical Society of America, 66(4), 1001-1017.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel contexts. Journal of the Acoustical Society of America, 67, 648-662.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. i. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. Journal of the Acoustical Society of America, 101(4), 2299-2310.

Broen, P. A., Strange, W., Doyle, S. S., & Heller, J. H. (1983). Perception and production of approximant consonants by normal and articulation-delayed preschool children. Journal of Speech and Hearing Research, 28, 601-608.

Brown, J. W. (1983). Stimulation maps from the standpoint of aphasia study: Commentary to Ojemann's *Brain organization for language from the perspective of electrical stimulation mapping*. Behavioral & Brain Sciences, 6(2), 207-208.

Chistovich, L. A., & Lublinskaya, V. V. (1979). The 'center of gravity effect in vowel spectra and critical distance between the formants: Psychoacoustic study of the perception of vowel-like stimuli. Hearing Research, 1, 185-195.

Chistovich, L. A., Sheikin, R. L., & Lublinskaja, V. V. (1979). "Centres of gravity" and spectral peaks as the determinants of vowel quality. In B. Lindblom & S. Öhman (Eds.), Frontiers of Speech Communication Research. Festschrift for Gunnar Fant. (pp. 143-157) London: Academic Press.

Churchland, P. S. (1983). Ojemann's data: Provocative but mysterious: Commentary to Ojemann's *Brain organization for language from the perspective of electrical stimulation mapping*. Behavioral & Brain Sciences, 6(2), 211-212.

Cohen, J., & Cohen, P. (1983). Applied multivariate regression/correlation analysis for the behavioral sciences. (2nd Edition ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. H., & Diehl, C. F. (1963). Relation of speech-sound discrimination ability to articulation-type speech defects. Journal of Speech and Hearing Disorders, 28, 187-190.

Cooper, W. E. (1974). Perceptual-motor adaptation to a speech feature. Perception and Psychophysics, 16(2), 229-234.

Cooper, W. E. (1983). Brain cartography: Electrical stimulation of processing sites of transmission lines? Commentary to Ojemann's *Brain organization for language from the perspective of electrical stimulation mapping*. Behavioral & Brain Sciences, 6(2), 212-213.

Cooper, W. E., Billings, D., & Cole, R. A. (1976). Articulatory effects on speech perception: A second report. Journal of Phonetics, 4(3), 219-232.

Cooper, W. E., Blumstein, S. E., & Nigro, G. (1975). Articulatory effects on speech perception: A preliminary report. Journal of Phonetics, 3(2), 87-98.

Cooper, W. E., Ebert, R. R., & Cole, R. A. (1976). Speech perception and production of the consonant cluster [st]. Journal of

Experimental Psychology: Human Perception and Performance, 2(1), 105-114.

Cooper, W. E., & Lauritsen, M. R. (1974). Feature processing in the perception and production of speech. Nature, 252, 121-123.

Cooper, W. E., & Nager, R. M. (1975). Perceptuo-motor adaptation to speech: an analysis of bisyllabic utterances and a neural model. Journal of the Acoustical Society of America, 58(1), 256-265.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of America, 27(4), 769-733.

Diehl, R. L. (1981). Feature detectors for speech: A critical reappraisal. Psychological Bulletin, 89, 1-18.

Diehl, R. L., Kluender, K. R., & Parker, L. M. (1985). Are selective adaptation and contrast effects really distinct? Journal of Experimental Psychology: Human Perception and Performance, 11, 209-220.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context dependent cues. Perception & Psychophysics, 22(2), 109-122.

Eimas, P. D., Cooper, W. E., & Corbit, J. D. (1973). Some properties of linguistic feature detectors. Perception & Psychophysics, 13(2), 247-252.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. Cognitive psychology, 4, 99-109.

Elman, J. L. (1979). Perceptual origins of the phoneme boundary effect and selective adaptation to speech: A signal detection theory analysis. Journal of the Acoustical Society of America, 65, 190-207.

Elman, J. L., & McClelland, J. L. (1986). Exploiting lawful variability in the speech wave. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech processes. (pp. 360-380) Hillsdale, NJ: Lawrence Erlbaum.

Fischer-Jørgensen. (1954). Acoustic analysis of stop consonants. Miscellanea Phonetica, II, 42-59.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. Journal of the Acoustical Society of America, 93(3), 1589-1608.

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. Phonetica, 43, 155-171.

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. Journal of Phonetics, 15(1), 67-83.

Flege, J. E., & Schmidt, A. M. (1995). Native speakers of Spanish show rate-dependent processing of English stop consonants. Phonetica, 52, 90-111.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. Journal of the Acoustical Society of America, 84, 115-123.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist approach. Journal of Phonetics, 14, 3-28.

Fowler, C. A. (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. Perception & Psychophysics, 55(6), 597-610.

Fox, R. A. (1978). Individual perceptual variation and a perception/production link in vowels. Papers from the 14th Regional Meeting, Chicago Linguistic Society, 98-107.

Fox, R. A. (1982). Individual variation in the perception of vowels: Implications for a perception-production link. Phonetica, 39, 1-22.

Frazier, L. (1983). Motor theory of speech perception or acoustic theory of speech production? Commentary to Ojemann's *Brain* *organization for language from the perspective of electrical stimulation mapping*. Behavioral & Brain Sciences, 6(2), 213-214.

Frieda, E. M. (1997, June 16-20). Comparison of native English speakers' perception and production of the English vowel /i/. Paper presented at the 133rd meeting of the Acoustical Society of America, State College, PA.

Ganong, W. F., III. (1978). The selective adaptation effects of burst-cued stops. Perception & Psychophysics, 24, 71-83.

Garrison, L. F., & Sawusch, J. R. (1986). Adaptation of place perception for stops: Effects of spectral match between adaptor and test series. Perception & Psychophysics, 40, 419-430.

Goldinger, S. D. (1997). Words and voices: Perception and production in an episodic lexicon. In K. Johnson & J. W. Mullennix (Eds.), Talker variability in speech processing. (pp. 33-66) San Diego: Academic.

Griffiths, S. K., & Johnson, C. J. (1995). Effects of training on fricative identification in toddlers. Applied Psycholinguistics, 16, 443-462.

Groenen, P., Maasen, B., Crul, T., & Thoonen, G. (1996). The specific relation between perception and production errors for place of articulation in developmental apraxia of speech. Journal of Speech and Hearing Research, 39(3), 468-482.

Haggard, M. P., Corrigall, J. M., & Legg, A. G. (1971). Perceptual factors in auditory defects. Folia Phoniatrica, 23, 33-40.

Halle. (1964). On the basis of phonology. In J. Fodor & J. Katz (Eds.), The structure of language. Englewood cliffs.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech, 1(1), 1-7.

Hazan, V., & Rosen, S. (1991). Individual variability in the perception of cues to place contrasts in initial stops. Perception and Psychophysics, 49(2), 187-200.

Hedrick, M. S., & Ohde, R. N. (1993). Effect of relative amplitude of frication on perception of place of articulation. Journal of the Acoustical Society of America, 94(4), 2005-2026.

Heinz, J. M., & Stevens, K. N. (1961). On the properties of voiceless fricative consonants. Journal of the Acoustical Society of America, 33(5), 589-596.

Hoffman, P. R., Daniloff, R. G., Alfonso, P. J., & Schuckers, G. H. (1984). Multiple-phoneme-misarticulating children's perception and production of voice onset time. Perceptual and Motor Skills, 58, 603-610.

Hoffman, P. R., Daniloff, R. G., Bengoa, D., & Schuckers, G. H. (1985). Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. Journal of Speech and Hearing Disorders, 30, 46-53.

Hoffman, P. R., Stager, S., & Daniloff, R. (1983). Perception and production of misarticulated /r/. Journal of Speech and Hearing Disorders, 48, 210-215.

Hoit-Dalgaard, J., Murray, T., & Kopp, H. G. (1983). Voice onset time production and perception in apraxic subjects. Brain and Language, 20(2), 329-339.

Jamieson, D. G., & Cheesman, M. R. (1986). Locus of selective adaptation in speech perception. Journal of Experimental Psychology: Human Perception and Performance, 12, 286-294.

Jamieson, D. G., & Rvachew, S. (1992). Remediating speech production errors with sound identification training. Journal of Speech-Language Pathology & Audiology, 16, 201-210.

Jassem, W. (1965). The formants of fricative consonants. Language and Speech, 8, 1-16.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: phonetic targets are hyperarticulated. Language, 69(3), 505-528.

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. Journal of the Acoustical Society of America, 94(2), 701-714.

Kent, R. D. (1983). Windows to the brain: Functional impairment and the surgical field. Commentary to Ojemann's *Brain organization for language from the perspective of electrical stimulation mapping.* Behavioral & Brain Sciences, 6(2), 214-215.

Kewley-Port, D. (1983). Time-varying features as correlations of place of articulation in stop consonants. Journal of the Acoustical Society of America, 73, 322-335.

Kewley-Port, D., & Luce, P. A. (1984). Time-varying features of initial stop consonants in auditory running spectra: A first report. Perception & Psychophysics, 35(4), 353-360.

Kewley-Port, D., Pisoni, D. B., & Studdert-Kennedy, M. (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. Journal of the Acoustical Society of America, 73, 1779-1793.

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

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America, 67, 971-995.

Kronvall, E. L., & Diehl, C. F. (1954). The relationship of auditory discrimination to articulatory defects of children with no known organic impairment. Journal of Speech and Hearing Disorders, 19, 335-338.

Ladefoged, P. (1982). A Course in Phonetics. (2nd ed.). San Diego: Harcourt Brace Jovanovich.

Lahiri, A., Gewirth, L., & Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-linguistic study. Journal of the Acoustical Society of America, 76, 391-404.

Lapko, L. L., & Bankson, N. W. (1975). Relationship between auditory discrimination, articulation stimulability, and consistency of misarticulation. Perceptual and Motor Skills, 40, 171-177.

Lehman, M. E., & Sharf, D. J. (1989). Perception/production relationships in the development of the vowel duration cue to final consonant voicing. Journal of Speech and Hearing Research, 32, 803-815.

Lewis, F. C. (1977). Distinctive feature confusions in production and discrimination of selected consonants. Language and Speech, 60-67.

Liberman, A. M., Cooper, F. S., Harris, K. S., & MacNeilage, P. F. (1962). A motor theory of speech perception. Speech Transmission Laboratory, Royal Institute of Technology, Stockholm.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74(6), 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech revised. Cognition, 21, 1-36.

Lindblom, B. (1963). On vowel reduction . Stockholm, Sweden: The Royal Institute of Technology, Speech Transmission Laboratory.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20(3), 384-422.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics, Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967. (pp. 563-567) Prague: Academia.

MacKay, I. R. A. (1987). Phonetics: The science of speech production. (2nd ed.). Boston: Little, Brown and Co.

MacNeilage, P. F., Rootes, T. P., & Chase, R. A. (1967). Speech production and perception in a patient with severe impairment of somesthetic perception and motor control. Journal of Speech and Hearing Research, 10, 449-467.

Mange, C. V. (1960). Relationships between selected auditory perceptual factors and articulation ability. Journal of Speech and Hearing Research, 3, 67-74.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. Perception & Psychophysics, 28(3), 213-228.

Marquardt, T. F., & Saxman, J. H. (1972). Language comprehension and auditory discrimination in articulation deficient kindergarten children. Journal of Speech and Hearing Research, 15, 382-389.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive psychology, 18, 1-86.

McClelland, J. L., & Rumelhart, D. E. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. (Vol. 2: Psychological and biological models). Cambridge, MA: MIT Press.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. Perception & Psychophysics, 46, 505-512.

Monnin, L. M., & Huntington, D. A. (1974). Relationship of articulatory defects to speech-sound identification. Journal of Speech and Hearing Research, 17, 352-366.

Nearey, T. M. (1992). Context effects in a double-weak theory of speech perception. Language and Speech, 35(1,2), 53-171.

Ojemann, G., & Mateer, C. (1979). Human language cortex: Localization of memory, syntax, and sequential motor-phoneme identification systems. Science, 205, 1401-1403.

Ojemann, G. A. (1983). Brain organization for language from the perspective of electrical stimulation mapping. Behavioral & Brain Sciences, 6(2), 189-230.

Paliwal, K. K., Lindsay, D., & Ainsworth, W. A. (1983). Correlation between production and perception of English vowels. Journal of Phonetics, 11, 77-83.

Patterson, R. D. (1974). Auditory filter shape. Journal of the Acoustical Society of America, 55(4), 802-809.

Perkell, J. H., & Nelson, W. L. (1985). Variability in production of the vowels /i/ and /a/. Journal of the Acoustical Society of America, 77(5), 1889-1895.

Perkell, J. S., & Matthies, M. L. (1992). Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability. Journal of the Acoustical Society of America, 91(5), 2911-2925.

Pickett, J. M. (1980). The sounds of speech communication: A primer of acoustic phonetics and speech perception. Baltimore: University Park Press.

Prins, D. (1963). Relations among specific articulatory deviations and responses to a clinical measure of sound discrimination ability. Journal of Speech and Hearing Disorders, 28, 382-288.

Raaymakers, E. M. J. A., & Crul, T. A. M. (1988). Perception and production of the final /s-ts/ contrast in dutch by misarticulating children. Journal of Speech and Hearing Disorders, 53, 262-270.

Repp, B. H. (1981). Two strategies in fricative discrimination. Perception & Psychophysics, 30(3), 217-227.

Richardson, K. H. (1992). An analysis of invariance in English stop consonants. Dissertation Abstracts International, 53(3-B), 1633.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. Perception & Psychophysics, 30, 309-314.

Rumelhart, D. E., & McClelland, J. L. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. (Vol. 1: Foundations). Cambridge, MA: MIT Press.

Rvachew, S., & Jamieson, D. G. (1989). Perception of voiceless fricatives by children with a functional articulation disorder. Journal of Speech and Hearing Disorders, 54, 193-208.

Samuel, A. (1986). Red herring detectors and speech perception: In defense of selective adaptation. Cognitive Psychology, 18, 452-499.

Samuel, A. (1988). Central and peripheral representation of whispered and voiced speech. Journal of Experimental Psychology: Human Perception and Performance, 14, 379-388.

Samuel, A. G. (1989). Insights from a failure of selective adaptation: Syllable-initial and syllable-final consonants are different. .

Samuel, A. G., Kat, D., & Tartter, V. C. (1984). Which syllable does an intervocalic stop belong to? A selective adaptation study. Journal of the Acoustical Society of America, 76.

Sander, E. K. (1972). When are speech sounds learned? Journal of Speech and Hearing Disorders, 37(1), 55-63.

Sawusch, J. R. (1976). Selective adaptation effects on end-point stimuli in a speech series. Perception & Psychophysics, 20, 61- 65.

Sawusch, J. R. (1977). Peripheral and central processing in selective adaptation of place of articulation in stop consonants. Journal of the Acoustical Society of America, 62, 738-750.

Sawusch, J. R. (1988). A computational approach to speech perception: An analysis of the role of auditory metrics in perception (Speech Research Status Report #3). Buffalo, NY: State University of New York at Buffalo.

Sawusch, J. R. (1996). Instrumentation and methodology for the study of speech perception. In N. J. Lass (Ed.), Principles of Experimental Phonetics. (pp. 525-550) St. Louis, MO: Mosby.

Sawusch, J. R., & Dutton, D. L. (1992). Computational metrics for place of articulation information in consonants and vowels. Speech Research Status Report, 4, 3-33.

Sawusch, J. R., & Jusczyk, P. (1981). Adaptation and contrast in the perception of voicing. Journal of Experimental Psychology: Human Perception and Performance, 7, 408-425.

Sawusch, J. R., & Pisoni, D. B. (1978). Simple and contingent adaptation effects for place of articulation in stop consonants. Perception & Psychophysics, 23, 125-131.

Scharf, B. (1970). Critical Bands. In J. V. Tobias (Ed.), Foundations of Modern Auditory Theory. New York: Academic.

Schmidt, A. M., & Flege, J. E. (1995). Effects of speaking rate changes on native and nonnative speech production. Phonetica, 52, 41-54.

Schmidt, A. M., & Flege, J. E. (1996). Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/English bilinguals. Phonetica, 53(3), 162- 179.

Sherman, D., & Geith, A. (1967). Speech sound discrimination and articulation skill. Journal of Speech and Hearing Research, 10, 277-280.

Shuster, L. I. (1990). Motor-motor adaptation to speech: Further investigations. Perceptual & Motor Skills, 71(1), 275-280.

Shuster, L. I., & Fox, R. A. (1989). Motor-motor adaptation: Preliminary findings. Perceptual and motor skills, 69, 435-441.

Simon, H. J., & Studdert-Kennedy, M. (1978). Selective anchoring and adaptation of phonetic and nonphonetic continua. Journal of the Acoustic Society of America, 64, 1338-1357.

Smith, B. L. (1992). Relationships between duration and temporal variability in children's speech. Journal of the Acoustical Society of America, 91(4), 2165-2174.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. Journal of the Acoustical Society of America, 70(4), 976-984.

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.

Stitt, C. L., & Huntington, D. A. (1969). Some relationships among articulation, auditory abilities, and certain other variables. Journal of Speech and Hearing Research, 12, 576-593.

Strange, W., & Broen, P. A. (1981). The relationship between perception and production of /w/, /r/, and /l/ by three-year-old children. Journal of Experimental Child Psychology, 31, 81-102.

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

Studdert-Kennedy, M. (1983). Mapping speech: More analysis, less synthesis, please. Commentary to Ojemann's *Brain organization for language from the perspective of electrical stimulation mapping.* Behavioral & Brain Sciences, 6(2), 218-219.

Summerfield, Q., Bailey, P. J., & Erickson, D. (1980). A note on perceptuo-motor adaptation of speech. Journal of Phonetics, 8(4), 491-499.

Sussman, H. M. (1989). Neural coding of relational invariances in speech: Human language analogs to the barn owl. Psychological Review, 96(4), 631-642.

Sussman, H. M. (1991). The representation of stop consonants in three-dimensional space. Phonetica, 48, 18-31.

Sussman, H. M., Hoemeke, K. A., & Ahmed, F. S. (1993). A cross-linguistic investigation of locus equations as a phonetic discriptor for place of articulation. Journal of the Acoustical Society of America, 94(3), 1256-1268.

Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. Journal of the Acoustical Society of America, 90(3), 1309-1325.

Syrdal, A. K. (1996, Oct. 3-6). Acoustic variability in spontaneous conversational speech of American English talkers. Paper presented at the 4th International Conference on Spoken Language Processing (ICSLP), Philadelphia.

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

Tomiak, G. R. (1991). An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents. Dissertation Abstracts International, 51(8-B), 4082-4083.

Travis, L. E., & Rasmus, B. (1931). The speech sound discrimination ability of cases with functional disorders of articulation. The Quarterly Journal of Speech, 17, 217-226.

Waldman, F. R., Singh, S., & Hayden, M. E. (1978). A comparison of speech-sound production and discrimination in children with functional articulation disorders. Language and Speech, 21, 205-220.

Walley, A. C., & Carrell, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. Journal of the Acoustical Society of America, 73, 1011-1022.

Weiner, F. F., & Falk, M. L. (1972). Speech-sound discrimination skills as measured by reaction time for normal and articulatory defective speaking children. Perceptual and Motor Skills, 34, 595-600.

Weiner, P. S. (1967). Auditory discrimination and articulation. Journal of Speech and Hearing Disorders, 32, 19-28.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[ʃ] boundary. Journal of the Acoustical Society of America, 69(1), 275-282.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. Perception & Psychophysics, 35(1), 49-64.

Whalen, D. H. (1991). Perception of the English /s/-/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. Journal of the Acoustical Society of America, 90(4), 1776-1785.

Woolf, G., & Pilberg, R. (1971). A comparison of three tests of auditory discrimination and their relationship to performance on a deep test of articulation. Journal of Communication Disorders, 3, 239-249.

Yamada, R. A., & Tokhura, Y. i. (1990, November 18-22, 1990). Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese. Paper presented at the ICSLP 90, Kobe, Japan.

Zlatin, M. A. (1974). Voicing contrast: Perceptual and productive voice onset time characteristics of adults. Journal of the Acoustical Society of America, 56, 981-994.

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. Journal of the Acoustical Society of America, 68(5), 1523-1525.