Importance of Formants 1 and 2 in Consonant Perception

THE IMPORTANCE OF THE FIRST TWO FORMANTS IN THE PERCEPTION OF CONSONANTS

by

Andrée Durieux Smith

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

School of Human Communication Disorders
McGill University
Montreal

August 1974

© Andree Durieux Smith 1975

ABSTRACT

Andrée Durieux Smith

The Importance of the First Two Formants in the Perception of Consonants

Ph.D. degree

School of Human Communication Disorders

McGill University

The intelligibility and perception of place, manner and voicing were studied in 17 consonants with the vowels /I/ and /u/ in filtered CV and VC syllables containing: 1) the first and second formants at a comfortable listening level; 2) the second formant alone; 3) the first formant alone; and 4) both formants at a high sound pressure level. Results indicated: 1) voicing and manner are well perceived when both formants are present at a comfortable listening level, but place errors still occur; 2) the second formant contains important place, manner and voicing information; 3) the first formant contains much less information, being primarily useful for voicing distinctions, with some manner information for plosives; 4) high sound pressure level slightly decreases the accuracy with which place is perceived; 5) intelligibility and perception of the three features vary as a function of the vowel environment and the positions of the consonants; and 6) invariant energy and transitions carry important information for consonant perception.

RESUME

Andrée Durieux Smith

L'Importance des Deux Premiers Formants pour la Perception des Consonnes

Ph.D. degree
School of Human Communication Disorders
McGill University

L'intelligibilité et la perception du lieu d'articulation, du mode de production et du voisement furent étudiés dans 17 consonnes avec les voyelles /I/ et /u/ dans des syllabes CV et VC filtrées contenant: 1) le premier et le deuxième formant à un niveau d'entente confortable; 2) seulement le deuxième formant; 3) seulement le premier formant; 4) les deux formants à un niveau de haute pression sonore. Les résultats indiquèrent que: 1) le voisement et le mode sont bien perçus lorsque les deux formants sont présents à un niveau d'entente confortable, mais qu'il y a encore des erreurs de lieu; 2) le deuxième formant contient des renseignements importants en ce qui concerne le lieu, le mode et le voisement; 3) le premier formant contient beaucoup moins de renseignements mais est utile pour faire les distinctions vocales, et les distinctions de mode pour les plosives; 4) le niveau de haute pression sonore augmente légèrement le nombre d'erreurs de lieu; 5) l'intelligibilité et la perception des trois traits varient en fonction du contexte vocalique et des positions des consonnes; et que 6) l'énergie invariable et les transitions contiennent des renseignements importants pour la perception des consonnes.

ACKNOWLEDGEMENTS

This research was supported by Canadian Medical Research Council Grant MA-1652 awarded to Dr. D. G. Doehring.

I am grateful to my supervisor Dr. K. K. Charan for his help in planning the study and also for his technical advice and assistance.

I am deeply indebted to Dr. D. G. Doehring for his invaluable suggestions and guidance in the areas of both organization and presentation of the material. The editorial comments of Dr. D. Ling and Dr. B. Bartholomeus are very much appreciated, as are the suggestions of Dr. M. Corballis and Dr. J. Ramsay with regard to the statistical treatment of the data. Special thanks are due to Sally Naish who typed the manuscript, prepared the figures and helped in the verification of the text.

I would also like to thank my parents who have always encouraged me in pursuing my studies and finally, my husband, for without his assistance and moral support this project would never have been completed.

TABLE OF CONTENTS

INTRODUCTION

Consonant sounds are produced by rapid and precise adjustments of the vocal tract. The dimensions of adjustment essential for their production have been described as distinctive features (Jakobson and Halle, 1956), with the assumption that correlates of these features are to be found at every level of the speech process (articulatory, acoustic, auditory). Descriptions of the acoustic correlates of distinctive features have been refined to the extent where they no longer provide a sufficient basis for practical application (Fant, 1962a). A more practical approach, which is applicable to both natural and synthetic speech, consists of using the broader features of place, manner and voicing, which appear to be essential for consonant identification. Place of articulation refers to the area of major constriction in the vocal tract, such as labial, alveolar, interdental, labiodental, palatal and velar. Manner of production specifies the type of articulation, such as plosive, fricative, nasal and liquid. The voiced-voiceless opposition indicates whether the vocal cords vibrate when the consonant is produced. It has been shown that both confusions (Miller and Nicely, 1955; Pickett, 1968) and judgements of similarity among consonants (Denes, 1963; Peters, 1963) can be adequately explained in terms of these three dimensions.

Acoustic analysis of speech sounds shows that the majority of English consonants coarticulated with vowels consist of short periods of relatively invariant energy within certain broad frequency bands, as well as periods of rapid changes in relatively narrower frequency bands. These concentrations of energy in certain bands of frequency are known as formants and are the result of the resonating properties of the vocal tract. Although the term 'formant' is mostly used in conjunction with vowel sounds, formant structures are also apparent in the invariant part of the spectrum of consonants (Jassem, 1965; Fant, 1968). The invariant energy of consonants is the result of the breath stream which is impeded by the closures produced by the articulators, and the rapid frequency changes or formant transitions occur as a natural consequence of the coarticulation of one phoneme with another and vary for a given consonant in different vowel environments.

One view of speech perception, based on research with synthetic speech, emphasizes the importance of transitions and more particularly second formant transitions, as cues for place of articulation and suggests that the second formant transition is probably the single most important carrier of linguistic information (Liberman, 1957; 1970; Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967). However, the results of studies with natural speech indicate that other aspects of the speech wave, such as invariant energy (Fant, 1967a; Cole and Scott, 1972; 1973a) and overall spectrum shape (Fant, 1968), are more important for consonant perception than band-limited cues, such as single formant transitions.

There have been as yet no studies involving natural or synthetic speech that have systematically assessed the contribution of single formants to consonant intelligibility. This was accomplished in the present study by band-pass filtering the consonants /p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, θ, ð, m, n, l/ in CV and VC syllables of natural speech in such a way that cues to consonant perception contributed by the first formant alone and by the second formant alone could be compared with the intelligibility of consonants when both formants were present. The results were analyzed not only in terms of overall intelligibility, but also with respect to the amount of place, manner and voicing information contributed by each formant. As the filtering was based on the first and second formant frequency regions, the more broad-band information from invariant energy, such as bursts for plosives and friction noises for fricatives, was reduced. In the case of nasals and liquids which have resonance characteristics analogous to those of the vowels, the spectrum reduction caused by the filtering was similar to that occurring for the vowels.

The results of this study contribute to knowledge concerning the role of acoustic cues in speech perception and to the resolution of some of the seemingly paradoxical disparities between acoustical characteristics of the speech wave and the identification of consonant sounds.

The following is a review of the literature relevant to the choice of parameters in the present study. The first section reviews the indirect evidence for the importance of the first and second formants in natural speech as seen from the results of amplitude and frequency distortion studies. The second section reports the results of studies dealing with the effects of distortion on place, manner and voicing distinctions. In the third section, the important cues for the identification of the features of place, manner and voicing for the consonants used in the present study are reviewed. Finally, the last section shows evidence for the possible masking of second formant information by the first formant under adverse listening conditions such as high sound pressure levels.

REVIEW OF RELEVANT LITERATURE

Vibration of the vocal cords results in the production of a complex sound characterized by a fundamental frequency which corresponds to the frequency of vibration of the vocal folds, and by a related harmonic structure. The vocal tract, which has variable resonant frequencies, emphasizes the harmonics of the vocal cord wave at a number of different frequencies (Denes and Pinson, 1963). The frequencies of vocal tract resonances are called formants and range from 270-739 Hz for the first formant, 840-2290 Hz for the second formant and 1690-3010 Hz for the third formant, for male adult speakers (Peterson and Barney, 1952).

Indirect Evidence for the Importance of the First and the Second Formant

Early research on the information-bearing parts of speech was largely influenced by the demands of telephonic communication; its aim was to estimate how much distortion could be imposed on the speech wave by frequency bandwidth reduction (Fletcher, 1929; French and Steinberg, 1947) or amplitude peak clipping (Licklider, 1946; Licklider and Pollack, 1948; Licklider and Miller, 1960) without reducing its intelligibility.

Frequency distortion studies

The relative contribution of frequency bands to speech intelligibility has been studied by determining the loss of intelligibility occurring when relatively large portions of the frequency spectrum are removed by filtering, with low frequencies eliminated by high-pass filtering, high frequencies eliminated by low-pass filtering, and both selected low and high frequencies eliminated by band-pass filtering. An important characteristic of filters is the rejection slope which refers to the rate of attenuation of frequencies beyond the filter cut-off points.

French and Steinberg (1947) determined the intelligibility of nonsense syllables as a function of the cut-off frequencies of high-pass and low-pass filters for adult speech. The intelligibility of the received speech sounds was measured in terms of a quantity called the articulation score which varied between 0 and 100%. An abrupt increase in intelligibility was observed as more of the second formant range was included by the filter when approached from either the low or high frequency direction. French and Steinberg divided the total range of speech frequencies (250-7000 Hz) into 20 contiguous frequency bands each of which contributed 5% to the articulation index. A total of ten bands fell within the second formant range. With respect to the low frequencies characteristic of first formants, the virtual elimination of low frequency information by variation of the filter cut-off frequency up to 800 Hz caused less than 10% decrease in articulation score. Similar results had been obtained by Fletcher (1929), who high-pass filtered speech at 1000 Hz and obtained an 80% reduction of speech 'power' and only a 10% reduction of the articulation score. Results of this nature led to the classical description of the relative contribution of intelligibility of the various parts of the speech spectrum which designates the high frequencies as information-bearing and the low frequencies as the energy-bearing portions of the speech signal.

In an early study of band-pass filtering, Egan and Wiener (1946) used a single band-pass filter to determine articulation scores for nonsense syllables spoken by a male and in the presence of a masking noise 40 dB less than the average speech level. Since 70% of the nonsense syllables were correctly perceived when the band-pass covered the frequencies 870-2500 Hz, Egan and Wiener concluded that the frequencies of the speech in the 'central range' were more important for intelligibility than the low or high frequencies. The frequency range from 870-2500 Hz covers most of the second formant range for male speakers.

A later study by Kryter (1960) involved the use of as many as three band-pass filters each with a bandwidth of 500 Hz and a rejection slope of 70 dB/octave. The highest articulation score for any single band-pass was obtained with the 1250-1750 Hz band, confirming the conclusions of Egan and Wiener which indicated that the frequencies contributing most to intelligibility were in the second formant range. When Kryter presented speech in two band-passes, the highest scores were achieved with a centre frequency on one band at 500 or 750 Hz and the centre frequency of the other band-pass either in the region 1500-1750 Hz or 2500-2750 Hz. In addition to contributing greatly to the intelligibility of the speech signal, the lower bands also made the speech sound natural, giving it a 'voiced quality'. These results indicate that the low frequencies of first formants do contribute to intelligibility.

Similar results have been found in other studies presenting speech in two filtered bands, many of which have involved attempts to develop audiometric tests for the diagnosis of lesions of the central auditory pathways (Matzker, 1962; Linden, 1964; Palva, 1965; Hayashi, Ohta and Morimoto, 1966). The aim of these tests was to present speech in a combination of two bands and to compare the results with the discrimination obtained for each band separately. In a normal subject the discrimination value for the two-band condition exceeded the arithmetic sum of the discrimination scores for each band. Before selecting the band-passes for his test, Palva

(1965) investigated band-pass combinations taken from the frequency areas between 400 and 1000 Hz and between 1200 and 3000 Hz. He found that the frequencies most important for discrimination of phonetically balanced Finnish words were contained within the regions 500-700 Hz and 1300-2000 Hz. Bands selected from these two frequency regions also provided the best summation effect. Palva concluded that the summation depended mainly on the relationship between the formant structure of the words and the frequencies of the bands employed, and that the best summation was obtained when the frequencies in a pair of bands coincided with the first and second vowel formants.

Franklin (1969a; 1969b) studied the effect of a low frequency band of speech on consonant discrimination by normal-hearing listeners using the Fairbanks Rhyme Test (Fairbanks, 1958). A high band of 1020-2040 Hz (filter slope: 60-70 dB/octave) 'rich' in consonant information was presented at threshold and a low band of 240-480 Hz containing 'minimum consonant information' was presented at threshold and also at 20 and 40 dB sensation level. When the high band was presented alone, consonant discrimination was 40%. When the low band was added at 0 dB, 20 dB and 40 dB sensation level, discrimination scores were 55, 61 and 38%, respectively. Thus, the addition of the low band to the high band contributed to intelligibility when added at 0 and 20 dB sensation level, but did not have the same effect on each of the consonants of the study.

None of the band-pass filter studies involved a precise removal of formant information, as was done in the present study. However, the results of these studies do demonstrate the importance of higher frequencies in the region of the second formant. Although the low frequencies do not function efficiently as a bearer of information when presented alone, they do improve recognition when presented together with high frequencies.

Amplitude distortion studies

One kind of amplitude distortion is the infinite amplitude clipping of a speech signal so that the original speech wave is reduced to a rectangular wave whose axis crossings represent the last vestiges of the original waveform. Speech is highly resistant to this kind of distortion. Licklider and Pollack (1948) demonstrated that infinitely clipped speech is 52% understandable and that speech which is given a high frequency emphasis by filter networks prior to clipping is 92% discriminable. This latter form of clipped speech is termed differentiated. Subsequent to the experiments of Licklider and Pollack, Chang, Pihl and Wiren (1951) observed that the average rate of axis crossings for differentiated speech corresponded reasonably closely to the second formant frequency, while the rate of unfiltered speech prior to clipping could be identified with the first formant frequency. Based on these observations, Thomas (1966; 1968a) hypothesized that the high intelligibility of differentiated speech might be related to the second formant frequency information retained in the distortion process. As spectrograms of clipped or differentiated-clipped speech reveal that while one formant is clearly predominant, all of the other formants are present, Thomas proceeded to 'isolate' the first and second formants by band-pass filtering before the clipping operation. Using a band-pass filter with an attenuation slope of 24 dB/octave, and maximum gain at the approximate logarithmic centre of 1500 Hz for the second formant extraction and of 500 Hz for the first formant extraction, Thomas filtered Egan's phonetically balanced words (Egan, 1948) and presented them to subjects for identification. The results for the clipped speech containing 'only second formant information', gave a 70.1% discrimination of the words. The intelligibility of clipped speech containing 'only first formant information' was 7.6%. Thomas concluded that speech in which all formants but the second have been suppressed is still highly intelligible. He attributed the intelligibility of differentiated-clipped speech to the survival of the second formant frequency information and hypothesized that the intelligibility of speech which has been subjected to any combination of amplitude, frequency or noise distortion is largely determined by the extent to which second formant frequency information survives the distortion process. However, the phonetically balanced words which were used in this study contain different vowels and therefore probably had a range of first and second formant frequencies. It seems unlikely that with band-pass filters with fixed cut-off frequencies and with a rejection slope of 24 dB/octave, listeners were being presented with 'only second formant information', and with 'only first formant information'. The 24 dB/octave rejection slope would only moderately attenuate the frequencies beyond the cut-off points of the band-pass.

As a more direct test of the major contribution of the second formant to intelligibility, Thomas (1966; 1968b) proceeded to generate synthetic speech from 'only' second formant information by using the extracted second formant information from natural speech to generate the synthetic speech signal. The same band-pass and clipper were used as in the previous experiment. Voice-excited synthetic speech resulted in an intelligibility of 64% for PB words. These results further confirmed that the second formant plays a 'predominant' role as a bearer of information, relating to the intelligibility of a speech signal.

Effects of Distortion on Place, Manner and Voicing Distinctions

Temporal segmentation

Sharf (1971) presented syllables containing initial fricatives and stops to listeners and had them rank the consonants on the perceptual parameters of duration, loudness, frequency, sharpness and contact. The scale values were compared on the basis of place, manner and voicing contrasts of articulation. The results indicated that duration is an appropriate perceptual parameter for the categorization of manner distinctions. Likewise, studies involving temporal truncations of syllables have demonstrated that manner is the phonetic feature least resistant to this kind of distortion (Gerstman, 1956; Barrs, 1963; Grimm, 1966; Cole and Scott, 1973b). Duration is necessary to distinguish among classes of sounds produced at the same place of articulation. The sequential elimination of segments of the initial part of a CV syllable containing a fricative will transform the consonant into an affricate and then into a stop (Gerstman, 1956; Barrs, 1963; Cole and Scott, 1973b). Temporal truncation has also been found to have an effect on voicing distinctions for initial plosives, voiceless plosives being heard as voiced with increased deletion of the initial part of syllables (Barrs, 1963; Grimm, 1966). Of the three phonetic features of place, manner and voicing, place is the most resistant feature to temporal segmentation (Grimm, 1966; Singh, 1966).

Random noise and frequency distortion

Miller and Nicely (1955) studied the perception of the phonetic features of voicing, nasality, affrication, duration and place of articulation in the initial consonants /p, b, t, d, k, g, f, v, s, z, θ, ð, ʃ, ʒ, m, n/ paired with the vowel /ɑ/, under conditions of random noise and low and high-pass filtering. According to the feature system used by Miller and Nicely, affrication distinguished the fricatives from the other consonants, as was the case of nasality for nasals; thus, both affrication and nasality represented the feature of manner of production. Duration further differentiated /s, z, ʃ, ʒ/ from all other stimuli. The results for the random noise conditions indicated that voicing and nasality were much less affected by noise than the other features, followed by affrication and duration which were somewhat more resistant than place. Both voicing and nasality were discriminable at a signal-to-noise ratio as poor as -12 dB while place was hard to distinguish at ratios less than 6 dB. The results for the low-pass conditions showed considerable correspondence with those of the random noise conditions, the most striking similarities being for place and voicing. The results for high-pass filtering indicated that all features deteriorated in about the same way. High-pass filtering removed most of the acoustic power of consonants, leaving them inaudible and consequently producing quite random confusions. The critical frequencies for the features studied were 450 Hz for nasality, 500 Hz for voicing, 750 Hz for affrication, 1900 Hz for place and 2200 Hz for duration.

Gay (1970), in a study on the intelligibility of the consonants /p, t, k, b, d, g, s, f, z, v, w, j, r, l, m, n/ in different vowel environments and under conditions of low-pass filtering, also found that place errors accounted for most of the confusions regardless of filter cut-off frequency, and that additional error types in the order of manner, voicing and nasality occurred at the lower cut-off frequencies

The results of these studies on the effect of distortions on the perception of the phonetic features of place, manner and voicing seem to indicate that time and frequency are the most important dimensions for manner and voicing distinctions while the parameters for place identification lie only along the frequency continuum. With respect to intensity, the place feature appears to be relatively weak as it is most affected by masking noise, while voicing is very resistant to this kind of distortion.

Important Perceptual Cues

In addition to the broader distortion studies described above which relate the phonetic features of place, manner and voicing to the dimensions of time, frequency and intensity, research has been carried out to identify the critical acoustical variables or cues (Haggard, 1972) responsible for particular distinctions. Three experimental techniques have been commonly employed in determining the important cues of speech: analysis of natural speech, production of synthetic speech, and finally, the splicing and gating out of segments of natural speech. Analysis of the spectrum of natural speech has flourished since the development of the spectrogram (Koenig, Dunn and Lacey, 1946), which provides a display of the speech signal along the dimensions of time, frequency and intensity, and has enabled the association of particular spectrographic patterns with particular phonemes (Potter, Kopp and Green, 1947). Synthesis has been an important tool for testing the relative importance of the various aspects of the sound pattern seen in the spectrogram (Fant, 1962a; Haggard, 1972). Much of the research on acoustic cues has been performed at Haskins Laboratories, using synthesizers of the pattern-playback type which reconvert the visual patterns of the spectrogram into sound after appropriate simplifications and modifications (Cooper, 1950; Borst, 1956). The assumption behind this method is that the spectrogram displays most prominently to the eye those features which are of greatest importance in auditory perception (Cooper, Liberman and Borst, 1951). There are several kinds of speech synthesizers (Cooper, 1964). Another approach involving vocal tract analog synthesizers (Stevens, Kasowski and Fant, 1953; Fischer-Jorgensen, 1961; Cooper, 1962; Fant, 1960; 1962a) has been used to study the relations between the articulatory and physical aspects of the sounds, but this method is not oriented towards the identification of the perceptually important parts of the speech spectrum. Research on the important acoustic cues with natural speech has mostly involved the splicing and recombining of recorded utterances and has been carried out to correct or corroborate the conclusions drawn from synthetic speech studies. The following is a review of studies on the important cues for plosives, fricatives, nasals and liquids, involving the three methods mentioned above.

Manner discrimination

In English, the manner feature is used more than any other articulatory criterion for distinguishing words from one another (Denes, 1963). In spite of the importance of manner distinctions, the acoustic correlates of manner cues have been least explored (Fant, 1967a).

Plosive manner cues. The invariant energy for plosives or stop consonants is characterized by a burst preceded by silence (Halle, Hughes and Radley, 1957). Rules for synthesizing plosives also usually call for a silent interval followed by a burst, although plosives can be synthesized without these parameters (Liberman, Ingemann, Lisker, Delattre and Cooper, 1959). Another important parameter for the synthesis of plosives is the first formant transition (Delattre, 1958; 1962). With the pattern-playback synthesizer, the best stop consonants are produced when the first formant starts at 240 Hz, the lowest frequency available on the synthesizer (Delattre, Liberman and Cooper, 1955). Stevens and House (1956) confirmed these findings by demonstrating, with a vocal tract analog, that when closure is made in the vocal tract, as in the case of a plosive, the lowest natural frequency of the system is 0 Hz.

The importance of the first formant for natural plosives was demonstrated by Halle, Hughes and Radley (1957). Listeners were presented with closed monosyllables containing the vowels /i, ɑ, u/ followed by each of the six stops and with open monosyllables containing the same vowels, but without final plosives. The syllables with the final stops were terminated after the vowel by means of an electronic gate, so that no burst appeared in the test. The subjects were asked to identify the end of the syllable. Halle et al.'s results agree with the results from Haskins Laboratories and indicate that the transition of the first formant plays a significant role in the identification of stops as a class. Furthermore, their results show that the extent of the transition is an important factor when considering the cue value of the first formant transition in natural speech. When the first formant was high and free to move downward, closed syllables were infrequently judged to be open. In contrast, syllables with low first formants were often judged to be open.

The importance of the speed of formant transitions as a cue for distinguishing stops from other classes ~~of phonemes~~ has been studied by Liberman, Delattre, Gerstman and Cooper (1956). The syllables /bɛ/ and /gɛ/ containing stops were synthesized and the duration of first and second formant transitions was varied in 10 msec. steps to determine whether changing transition speed was sufficient to convert the syllables into /wɛ/ and /jɛ/ containing semi-vowels, and then into vowels of changing colour, namely /uɛ/ and /iɛ/, as the speed of change further increased. The change from /b/ to /w/ occurred when the duration of the transitions reached 40 msec., while /g/ changed to /j/ around 50-60 msec. These results indicated that listeners were able to use the speed of first and second formant transitions as a cue for distinguishing between stop consonant, semi-vowel and vowel of changing colour. A second experiment separated the rate and duration aspects of transition speed. The distinction between /b/ and /w/ was studied in different vowel contexts. Plots of transition rate and duration against listener judgements indicated that duration rather than rate provided a sufficient cue for distinguishing between the two classes of phonemes.

Fricative manner cues. In natural speech, the invariant energy of fricatives is characterized by a friction noise which is the result of turbulence and other nonlinear processes in a vocal tract of complicated shape (Fant, 1960). Manner rules for the synthesis of fricatives call for an interval of band-limited noise (Liberman et al., 1959; Delattre, Liberman and Cooper, 1964) and for a first formant transition originating from a higher frequency than the stops (Delattre, 1958). The importance of transition duration has not been systematically studied in the differentiation of fricatives from other classes of consonants. In natural speech a secondary cue for fricatives may lie in the acoustical characteristics of adjacent vowels. Vowels in fricative environments are longer in duration, lower in fundamental frequency and greater in relative power than vowels in plosive environments (House and Fairbanks, 1953; Lintz and Sherman, 1961).

Nasal manner cues. Nasals are characterized by the presence of nasal resonances in the invariant part of the spectrum (Fant, 1968). For the synthesis of nasals, the intensity of nasal formants is somewhat less intense than those used to produce synthetic vowels (Liberman, Delattre, Cooper and Gerstman, 1954; Delattre, 1958). Experiments with vocal tract analogs (Pickett, 1965) and with nasals synthesized by an analysis-synthesis-computer scheme (Fujimura, 1962) indicate that the presence of nasal output is a very effective parameter for differentiating nasals from other classes. Fujimura found that the common characteristics of nasal murmurs were the existence of a very low first formant located around 300 Hz, well separated from the upper formant structure.

Manner rules for the synthesis of nasals also call for an absence of first formant transitions (Liberman et al., 1954). The speed of transitions appears to contribute to the plosive-nasal distinction. Pickett (1965) using a serial tract analog synthesizer found that the more gradual transition of nasals caused by the 'sluggish movement typical of velar articulation' was an important factor for the differentiation between /n/ and /d/.

A secondary cue for manner in final nasals is a nasalization of the adjacent vowel (Latif, Gallagher, Goldstein and Daniloff, 1971). One of the effects of the nasalization of a vowel is a damping (Delattre, 1954) and a subsequent increase in the bandwidth of the first formant (Nakata, 1959).

Liquid manner cues. Liquids in natural speech contain a vocalic like segment in the invariant part of the spectrum (Fant, 1968). This is represented in pattern-playback studies by steady-state formants which not only begin at the loci, but remain there 30-50 msec. before proceeding to the steady-state positions of the vowel (O'Connor, Gerstman, Liberman, Delattre and Cooper, 1957). Manner rules for liquids also call for a higher starting frequency of the first formant transition (360-400 Hz) than for any other class of consonants. In addition formants should have relatively slow transitions (75-100 msec.) drawn so as to be continuous with the locus formants (Delattre, 1958; Liberman et al., 1959). This continuity, according to Haskins' research, further distinguishes the liquids from the nasals, where the steady-state onsets typically differ in frequency from the starting point of the transitions (Cooper, Delattre, Liberman, Borst and Gerstman, 1952; O'Connor et al., 1957). This finding disagrees with Nakata's (1959), who successfully synthesized nasals with a terminal analog synthesizer in which the formants moved continuously from the consonant to the vowel segment. This discrepancy may be the result of differences in the synthesizers used in the two studies. Another factor which may play a role in the nasal-liquid distinction is the frequency position of the low formant of the resonant portion (Cooper et al., 1952).

In summary, manner cues are found in both the invariant and vocalic parts of the spectrum of consonants. With respect to the vocalic part, the first formant transition and the duration of transitions contribute to manner distinctions for some classes of consonants. The information content of single first formants with respect to the manner feature has not been assessed in natural speech.

Place discrimination

Research with synthetic speech has mostly investigated the importance of transitions, while natural speech studies have compared the place cues contained in the invariant and vocalic parts of the spectrum. Place cues have been more extensively studied in plosives than in any other class of consonants.

Plosive place cues.- Both transitions and bursts have been studied as place cues in plosives. The great majority of experiments on the importance of single transitions in stops have involved the use of pattern-playback synthesized speech. These experiments have systematically varied one parameter, such as second formant transitions (Cooper, Delattre, Liberman, Borst and Gerstman, 1952; Liberman, Delattre, Cooper and Gerstman, 1954; Delattre, Liberman and Cooper, 1955) and third formant transitions (Harris, Hoffman, Liberman, Delattre and Cooper, 1958), and have studied how these variations affect the perception of stops. The results of studies with second formant transitions added to two formant patterns of vowels indicate that the direction and degree of second formant transitions can serve as cues for the aurally perceived distinctions among both voiced and voiceless stops, subject agreement being greater for voiced stops. Rising transitions were generally heard as /p/ or /b/, and falling transitions were generally heard as /t, d/ or /k, g/ depending on the size of the transition and the vowel with which it was paired (Cooper et al., 1952, Liberman et al., 1954). Where second formants failed to achieve unanimous agreement among listeners as in the case of alveolars with some front vowels, third formants were seen to contribute to place discrimination (Harris et al., 1958). The pattern playback, however, cannot present isolated third formants in order to obtain a direct measure of the effect of this cue alone, as third formants must be presented with the first and second formant transitions in order to produce intelligible sounds. In pattern-playback studies, formant transitions were thought of as movements from an acoustic locus representing the place of articulation of a consonant to the steady-state level of the vowel and the perceptual importance of this locus required that transitions point, but not necessarily start, at the locus frequency (Delattre et al., 1955). Second formant locus frequencies were fixed for stops produced at the same place of articulation and were independent of vowel context. Results of experiments with vocal tract analogs agree with the second formant locus values found with the pattern playback for alveolars, but fail to confirm the single locus frequencies for labials and velars (Stevens and House, 1956). Similar irregularities have been noted in spectrographic studies of natural speech (Halle, Hughes and Radley, 1957; Lehiste and Peterson, 1961; Ohman, 1966; Fisher-Jorgensen, 1972c). The importance of transitions in natural speech has been studied in syllables containing unreleased final plosives (Andresen, 1960; Fisher-Jorgensen, 1972a) and final plosives in which the bursts have been gated out (Halle, Hughes and Radley, 1957). All these studies indicate that the value of all formant transitions as perceptual cues for consonants seems to depend on the magnitude of the formant movements which in turn depends on the articulatory position of the particular vowel and consonant involved. These observations have also been made by Wang (1959) and Wang and Fillmore (1961).

In addition to transitions, bursts have been studied as possible place cues in plosives. Liberman, Delattre and Cooper (1952) presented bursts synthesized with the pattern playback, to listeners for identification as /p, t/ or /k/. Bursts above 3240 Hz produced /t/ before any vowel whereas bursts at very low frequencies (360 Hz) produced /p/ before any vowel. Bursts in the middle frequencies produced either /p/ or /k/ depending on the following vowel. Liberman et al. concluded that although bursts may act as cues in ambiguous cases, the vowel played the critical part in the auditory perception of stops. Schatz (1954) confirmed these results for the burst portion of the phoneme /k/ using real speech. The burst from /k/ taken from syllables with /i, a/ and /u/ was heard as different stops when spliced with each of the vowels /i, a/ and /u/. With the exception of velars, Cole and Scott (1972; 1973a) showed that when bursts from initial stops were spliced onto the transitionless steady state of the vowel from which the burst was originally removed or of a different vowel, listeners were able to correctly identify the stops. These results demonstrated that the energy spectra which accompanies natural stop consonants in different vowel contexts contains invariant perceptual information. Similar results were also found by Malécot (1958) with final plosives. Malécot interchanged bursts and steady states of vowels containing transitions. His results indicated that bursts dominated everywhere even in cases where bursts were appended to first segments in which a different stop was involved. Malécot concluded that place cues contained in plosive bursts, and more particularly voiceless plosive bursts, were more powerful than those contained in the final transitions of the preceding vowel. Wang (1959), using a similar technique, also confirmed Malécot's results. With respect to voiceless plosives, Scheib and Winitz (1972) and Winitz, Scheib and Reeds (1972) found that the importance of the burst was in the order of /t, p/ and finally /k/, and that bursts isolated from conversational speech could be identified by listeners. Halle, Hughes and Radley (1957), Malécot (1958) and Cole and Scott (1972; 1973a) also obtained correct identifications of plosives from bursts isolated from syllables. These results indicate that bursts contain more powerful cues than the results of pattern-playback studies had suggested. The discrepancy may lie in the fact that synthetic bursts are not representative of natural speech. In natural speech, duration of the plosive burst and of aspiration vary for the different stops (Fisher-Jorgensen, 1954), and the frequency concentrations of the bursts cover a wider range than the 'teardrop' shape bursts of synthetic speech. Labials in natural speech were found to have a primary concentration in the low frequencies (500-1500 Hz), the alveolars were found to have a flat spectrum or one in which the higher frequencies (above 4000 Hz) predominated, and the velars showed strong concentrations of energy in intermediate frequency regions (Halle et al., 1957). Cues that make possible the identification of the bursts as different stops must reside in the spectrum. There has been no study investigating which parameters of the spectrum, such as overall spectrum shape or frequency regions corresponding to formants, are necessary for accurate perception.

As both bursts and transitions carry important information, some studies have been designed to determine the relative importance of these two cues. Experiments with pattern-playback synthesized speech have been unable to assess the relative contribution of second formant transition, third formant transition and burst in the perception of place, because of the strong influence of the second formant transition (Hoffman, 1958). Experiments with other speech synthesizers (Ainsworth, 1968a) and with natural speech (Fisher-Jorgensen, 1972b; 1972c) have indicated that the relative importance of various features depends on the following vowel. These studies have investigated the relative contribution of the overall pattern of transitions and of the burst, and have shown that when there is much tongue movement in coarticulation, the transitions are extensive and their perceptual importance greater. When there is little tongue movement, airflow is turbulent and a well defined and perceptually important noise burst is produced. Voiceless stops depend strongly on the noise bursts which play an important role in perception (Ainsworth, 1968a;

Fischer-Jorgensen, 1972d).

<u>Fricative place cues</u>. Sharf (1968) tested the distinctiveness of the place of articulation feature in fricatives by presenting an extensive range of fricatives to listeners for identification. His results indicate that listeners have considerable difficulty in identifying fricatives from auditory cues when the usual contextual cues are not available. Both transitions and friction noises have been studied as cues for place of articulation in fricatives.

Harris (1954; 1958) using natural speech, combined the noise portion from one spoken fricative vowel syllable with the vocalic portion from another fricative vowel syllable. Her results indicated that the important cues for the distinction between /s/ and /ʃ/ were given by the noise part of the syllable, but that the differentiation of /f/ and /θ/ was accomplished primarily on the basis of cues contained in the transition part of the syllable, whether from natural or pattern-playback synthetic speech (Harris, 1954). These results were confirmed by vocal tract analog (Heinz and Stevens, 1961) and natural speech studies (Cole and Scott, 1973b). Voiced fricatives were studied in a similar way (Harris, 1958) by recombining friction noises and vocalic portions from different syllables in natural speech. The results were similar to those for the voiceless fricatives, though not as clear. The phonemes /z/ and /ʒ/ behaved like their unvoiced counterparts. However, the results of /v/ and /ð/ were variable for different vowels and were thought to be attributed to the difficulty of splicing at the join of friction and vocalic portions. When the importance of the fricative noise is neutralised, formant transitions contribute to the identification of place in voiced fricatives. Delattre, Liberman and Cooper (1964) using the pattern-playback synthesizer found that both second and third formant transitions acted as cues for the voiced fricatives depending on vowel context, but that the cue value of formant transitions did not seem overwhelming.

With the exception of /f-θ/ and /v-ð/, fricatives are clearly identifiable when vowel transitions are removed from fricative vowel syllables (Harris, 1958; Heinz and Stevens, 1961; Cole and Scott, 1973b), and when friction noises are presented in isolation (Hughes and Halle, 1956; Strevens, 1960). These observations have led investigators to speculate that the important cues for the distinction between fricatives reside in the most prominant region of the friction noise. The filtering of synthetic voiceless fricatives (Harris, 1956) and the analysis of natural fricatives (Hughes and Halle, 1956; Strevens 1960; Jassem, 1965) has indicated that /s/ has strong concentrations of energy above 3600 Hz while /ʃ/ never has concentrations above this frequency. The energy for /f/ and /θ/ is concentrated in the mid frequencies. The relative levels of frequency peaks in the spectrum of natural fricatives can be considered as formants. Analysis of levels and frequencies indicates that /f, v, θ/ and /ð/ have their main concentration of energy in the second formant (/f,v/: 1550 Hz; /θ/: 1590 Hz; /ð/: 1460 Hz), that /ʃ/ and /ʒ/ are most concentrated in the third formant region (/ʃ/: 2610 Hz; /ʒ/: 2650 Hz), while /s/ and /z/ have concentrations in the fourth formant (/s/: 3950 Hz; /z/: 3920 Hz) (Jassem, 1965). In frequency regions greater than 1000 Hz, the spectra of voiced fricatives do not differ appreciably from those of the unvoiced (Hughes and Halle, 1956). There has been no study which has isolated the formant regions of natural fricatives in order to study their perceptual importance. In addition to concentrations of energy in various frequency regions, other spectral features, such as spectrum shape (Jassem, 1965), spectrum length (Strevens, 1960) and the intensity of the noise relative to the vowel (Harris, 1956; Hughes and Halle, 1956; Strevens, 1960), may play a role in the differentiation of the various fricatives.

Nasal place cues. Experiments with the pattern-playback synthesizer have indicated that second formant transitions were cues for the perceived distinctions among /m, n, ŋ/ (Cooper, Delattre, Liberman, Borst and Gerstman, 1952). Final nasals synthesized with neutral resonances were studied as cognates to both voiced and voiceless plosives. Rising formant transitions which had been heard as /b/ or /p/ were now heard as /m/. A comparable crossover in which /n/ paralleled /d/ or /t/, and /ŋ/ paralleled /g/ or /k/ was also seen. A comparison of the results of nasals with those of plosives indicated that second formant transitions were probably somewhat less effective as cues for nasals than for stops. Consonants articulated at the same place were found to have the same locus values for both classes of sounds. Experiments with terminal analog synthesized speech have confirmed the importance of second formant transitions for distinctions among nasals, but have found that second formant loci appropriate for the different nasals were not fixed, but somewhat dependent upon the adjacent vowel (Nakata, 1959). With respect to the starting frequencies of the second formant, vocal tract analog studies have shown that when consonants were identified as /m/, the second formant started between 800-1300 Hz, while the starting frequencies for /n/ were between 1500-2000 Hz and between 2000-2500 Hz for /ŋ/ (Hecker, 1962).

The importance of nasal resonances in isolation for the distinction between /m, n, ŋ/ has been studied with vocal tract analog synthetic speech (House, 1957; Nakata, 1959) and with natural speech (Malécot, 1956). Nakata using a vocal tract analog found that the maximum responses for /m/,

/n/ and /ŋ/ were 64, 52 and 41% respectively at frequencies 1100 Hz, 1700 Hz and 2300 Hz. Malécot (1956) using natural speech also found that /m/ was the most identifiable resonance, while /ŋ/ was the most difficult to identify. In general, synthetic resonances (House, 1957; Nakata, 1959) are more readily identifiable than natural ones (Malécot, 1956). This result may be explained by the fact that resonances are better defined in synthetic speech than in natural speech, because of the high damping caused by the soft walls and involved geometry of the natural nasal cavity. Both studies with natural and synthetic speech indicate that there is a tendency for listeners to identify all resonances as /m/. Lxperiments with natural CV and VC syllables in which transitions have been removed show results very similar to experiments in which resonances were presented in isolation (Malécot, 1956).

Studies which have sought to determine whether transitions or resonances were most important perceptually have shown that when resonances and transitions were interchanged in syllables containing nasals, stimuli containing an /m/ factor, whether transition or resonance, were heard as /m/. This was also the case for /n/, although to a lesser extent, while /ŋ/ required both transitions and resonances to be identified correctly (Malécot, 1956).

Liquid place cues. The place cues for liquids will not be reviewed as /l/ was the only liquid used in the present study. The liquid /r/ which is distinguished from /l/ by third formant transitions (O'Connor, Gerstman, Liberman, Delattre and Cooper, 1957; Lisker, 1957; Ainsworth, 1968b) was eliminated from the corpus of stimuli because of the inability of sufficiently attenuating unwanted formants by band-pass filtering.

Consonant position effects on place discrimination. Experiments with synthetic speech of the pattern-playback type represent final transitions as mirror images of initial ones and seem to indicate that the perceptual importance of transitions in both positions is the same (Liberman, Ingemann, Lisker, Delattre and Cooper, 1959). Analytic studies of natural speech reveal that transitions differ considerably depending upon the position of the consonant relative to the vowel (Stevens, House and Paul, 1966; Broad and Fertig, 1970), and that locus values for final consonants show great fluctuations (Lehiste and Peterson, 1961). Experiments with natural speech dealing with the identification of voiced plosives from syllables in which explosions have been removed (Fisher-Jorgensen, 1972c) and with the identification of voiceless plosives from bursts to which vowel segments were added (Winitz, Scheib and Reeds, 1972) indicate that transitions contribute more to the identification of final consonants. These results agree with the observation that place in natural speech is easier to perceive in the final, than in the initial position (Ahmed and Agrawal, 1969; Gupta, Agrawal and Ahmed, 1969).

Beiter and Sharf (1972) tested the identification of consonants from CV and VC vowel transitions from tapes played forward and backward. VC transitions were significantly better cues to consonant identification than CV transitions whether heard forward or backward.

Sharf and Hemeyer (1972) determined the relative importance of CV and VC formant transitions in the identification of place in the consonants /p, b, t, d, k, g, f, v, s, z, ʃ, ʒ/ paired with the vowel /ə/. The noise portions of the consonants were eliminated from the syllables, and listeners were asked to identify the formant transitions as labial, alveolar, palatal or no consonant. The results indicated that the number of correct identifications was considerably greater than chance for all sounds except voiceless stops, and that there were significantly more correct identifications made of VC than of CV transitions. The findings were explained on the basis of coarticulation. It was suggested that the effect of forward coarticulation on the assimilation of consonant features by vowels would make VC transitions more sufficient cues than CV transitions.

In summary, research with pattern-playback synthetic speech has indicated that the second formant transition is the most important carrier of place information for plosives, some fricatives and nasals. These results have been obtained with a technique which has kept all parameters constant and varied only second formant transitions. In addition, this method has neutralized the invariant parts of the spectrum such as bursts, frication noises and nasal resonances in order to fully study the effect of second formant transitions on perception. Research with natural speech is unable to use this approach and has concentrated on comparing place cues in the entire invariant part of the spectrum to cues in the entire vocalic (vowel steady state and transition) part. These comparisons show that there are more place cues in the invariant part of the spectrum than the results with pattern-playback synthetic speech indicate, and that the relative importance of vocalic and invariant parts is dependent on the extent of transitions as related to the coarticulation of the consonant with the adjacent vowel and to the consonant position relative to that vowel. Research with pattern-playback synthetic speech has been unable to assess how several cues share in the control of perception because of the very strong cue effect of the second formant. There has been no study with natural speech which has assessed the importance of single second formants or of band-limited frequency areas in the invariant part of the spectrum, relative to the place feature

Voicing discrimination

Plosive voicing cues. Voicing, aspiration and articulatory force have all been cited as a basis for separating the stop categories of natural English. The compelling reason for the selection of voicing as the most precise dimension for separating the two groups of stops is that aspiration has a contrastive value limited to particular contexts, while articulatory force has no agreed-upon physical correlate (Lisker and Abramson, 1967a). As seen in spectrograms, voicing yields harmonic excitation of a low frequency band known as voice bar, during the interval of oral closure. This voice bar is also a prerequisite for the synthesis of voiced stops (Liberman, Ingemann, Lisker, Delattre and Cooper, 1959). Because of the several distinctions between the groups /b, d, g/ and /p, t, k/, researchers have sought some single best measure by which to separate the two phoneme categories. It was discovered that the timing relation between voice onset and the release of occlusion could serve as a basis for separating the stop categories in various languages (Lisker and Abramson, 1964a; 1964b) and could lead to perceived differences in synthetic stops (Abramson and Lisker, 1965; Lisker and Abramson, 1967b). The choice of the voice onset time measure as an appropriate dimension for separating the two categories of stops does not mean that stops differ only with respect to the feature of pulsing. Voice onset time is the glottal mechanism which controls voicing and is responsible for generating aspiration and articulatory force (Fant, 1960). Aspiration is presumed to be related to the attenuation and cutback of the first formant.

Experiments with pattern-playback synthetic speech had indicated that strong voiced stops could only be synthesized by starting the first formant transition very low on the frequency scale (Cooper et al., 1952; Liberman, Delattre, Cooper and Gerstman, 1954; Delattre et al., 1955). This observation led to experiments which systematically eliminated a portion of the first formant transition and successfully transformed voiced stops into their voiceless cognates, indicating that the delay in the first formant was a sufficient cue for distinguishing between voiced and voiceless stops (Liberman, Delattre and Cooper, 1958). The delay in the first formant is a natural consequence of lengthening voice onset time. In addition, the substitution of noise in the second and third formants during the duration of the first formant cutback seemed to produce a greater impression of voicelessness than the first formant cutback alone. Noise in the region of second and third formants is an acoustic correlate of aspiration.

Both voice onset time and first formant cutback were studied simultaneously in experiments with synthetic speech (Stevens and Klatt, 1971; 1974; Haggard and Summerfield, 1972; Summerfield and Haggard, 1972). These experiments have manipulated both cues independently and have indicated that both cues share in the perception of voicing (Stevens and Klatt, 1971; 1974; Haggard and Summerfield, 1972). The presence of a first formant transition acts as a cue for voicing and a negative cue for voicelessness (Stevens and Klatt, 1971; Summerfield and Haggard, 1972; Haggard and Summerfield, 1972) while long voice onset times represent a major cue for voicelessness. A significant trading relationship exists between the two cues (Stevens and Klatt, 1974).

Fricative voicing cues. The important cues for voicing in fricatives have not been as systematically studied as in plosives. Some of the important parameters for voicing in these consonants can be drawn from the minimal rules necessary for their synthesis. The voicing rule for fricatives requires that the interval of band-limited noise necessary for manner, be of low intensity, have a duration of 100 msec. and be accompanied by a voice bar (Delattre, 1958; Liberman et al., 1959). In addition, low intensity formants are required to run through the friction (Delattre, Liberman and Cooper, 1964).

In the spectra of natural voiced fricatives, there exists a strong component in the region below 700 Hz, which is never prominent in voiceless fricatives. In the region greater than 1000 Hz, the spectra of voiced fricatives do not differ appreciably from those of the unvoiced fricatives. In addition, voiced fricatives are characterized by less hiss and less power than voiceless fricatives, since the breath stream is interrupted and reduced in flow by the action of the vocal cords (Hughes and Halle, 1956; Strevens, 1960).

General voicing cues. It appears from the results of Haggard, Ambler and Callow (1970), that changes in pitch at the onset of glottal vibration may distinguish voiced and voiceless stops for some listeners. Experiments on the identification of ambiguous syllables differing only in the pitch curve at voicing onset show that a low rising pitch leads to the perception of initial voiced stops, while a high falling pitch leads to the perception of voicelessness. Haggard et al. (1970) also expected pitch to assist the voicing distinction in some fricatives.

In addition to pitch, vowel duration preceding final voiced consonants has been found to be much longer than vowel durations preceding voiceless consonants in natural speech (House and Fairbanks, 1953; Denes, 1955; Halle, Hughes and Radley, 1957; Peterson and Lehiste, 1960; Lintz and Sherman, 1961; Broad and Fertig, 1970). On the other hand, the duration of voiceless consonants is much longer than the duration of voiced consonants. By recording the word /juz/ three times, and splicing the three /z/ endings onto the vowel, Denes (1955) was able to transform the word /juz/ into /jus/. Stimuli synthesized with the pattern playback, also showed that the duration of vowels and final consonants had a definite and consistent influence on the perception of voicing, but that these factors remained effective only in meaningful connected speech.

Raphael (1972) investigated the effect of varying preceding vowel duration upon the perception of word final stops, fricatives and clusters in synthetic speech. His results indicate that regardless of the cues for voicing or voicelessness used in the synthesis of the final consonant or cluster, listeners perceived the final segments as voiceless when they were preceded by vowels of short duration and as voiced when they were preceded by vowels of long duration. Duration was a more effective cue for stops and clusters than for fricatives.

Voiced consonants also have the effect of lowering the fundamental frequency and increasing the power of the adjacent vowel (House and Fairbanks, 1953).

In summary, the research evidence indicates that the important cues for voicing in plosives are the voice bar, first formant transition and the timing relation between voice onset and the release of occlusion which is accompanied by noise at the level of the second formant. For fricatives, the voicing cues have not been as extensively studied as in plosives, but lie in the voice bar, in the presence of a strong component in the invariant part of the spectrum and in the duration. When both classes of consonants are in the final position, vowel duration is an effective cue for voicing. There has been no study with natural speech which has assessed the contribution of single first formants to voicing.

The Low Frequencies as a Masker

The first formant of a vowel is always more intense than its second formant (Peterson and Barney, 1952; Fant, 1967b). In cases where listeners have a hearing loss it is necessary to present speech at a high sound pressure level. It has been shown that when complex sounds such as speech sounds are amplified, the high tones begin to be masked by the low tones (Fletcher, 1929; French and Steinberg, 1947; Palva, 1965). Similarly, when speech is presented in narrow bands, there is the possibility of one band of speech masking another (Kryter, 1960). House, Williams, Hecker and Kryter (1965) in the development and evaluation of a new speech intelligibility test also found that the speaker differences observed in their study could be tentatively explained by the masking of consonants by the vocalic portion of a word in one speaker.

French and Steinberg (1947) found that the two low frequency bands of 250-375 Hz and 375-505 Hz caused the greatest amount of interference with the higher bands. Similarly, Franklin (1969a; 1969b) found that a 240-480 Hz frequency band of speech acted as a masker for most consonants when presented at 40 dB SL to a 0 dB SL high band of speech. All three bands contain first formant frequencies.

According to Martin, Osberger and Pickett (1972), Martin, Pickett and Colten (1972), Danaher, Osberger and Pickett (1973) and Pickett and Danaher (1973), the upward spread of masking which occurs at high sound pressure levels in the presence of low frequency energy has been in part responsible for the reduced discrimination ability found in most subjects with sensorineural losses. Martin, Pickett and Colten (1972) studied the discrimination of synthetic formant transitions in listeners with severe sensorineural losses. Second formants were presented alone and in the presence of first formants. In the second formant alone condition, most subjects had low discrimination thresholds. However, when the subjects had to discriminate the frequency of the second formant in presence of the first, their discrimination was reduced and their thresholds were higher. When the first and second formants were close in frequency, the reduction in discrimination was great. The authors interpreted these findings as indicating that low frequency vowel formants can produce masking which reduces the ability to discriminate cues in the higher frequencies. When normal hearing listeners were tested at sound pressure levels of 95-105 dB, substantial amounts of masking also occurred and Martin, Pickett and Colten (1972) concluded that a reduced discrimination ability characterizes most subjects who listen at high sound pressure levels. Other variables were found to influence masking, namely the frequency region in which the second formant transitions occurred, the relative amplitudes of first and second formants, and finally the presence of a transition in the first formant (Danaher, Osberger and Pickett, 1973). First formant transitions were found to give the stimuli a more speech-like quality, but also to produce a masking of a different nature than that produced by the first formant steady state. These results could be related to the reduced discrimination that normally occurs within phoneme boundaries (Liberman, Harris, Hoffman and Griffith, 1957). First formants which had been successively cut back, as in voiceless stops, were also seen to produce a type of backward masking which could reduce the discrimination of frequency transitions which preceded them (Danaher and Pickett, 1973; Pickett and Danaher, 1973). Each of these masking effects was also found in normal hearing subjects who were tested at high sound pressure levels.

Methods to reduce the masking effect of the first formant have been studied by Pickett and Danaher (1973) who presented synthetic first and second formants dichotically and obtained improvements in second formant transition discrimination. A dichotic presentation of natural speech filtered in low and high frequency bands has also been found to decrease the masking effect and improve speech discrimination (Franklin, 1969a; 1969b; 1972).

Another approach to the problem of masking by low frequencies has been to reduce the energy of the first formant. Danaher, Osberger and Pickett (1973) found that synthetic second formant transition thresholds improved when the energy of the first formant was reduced. The authors did not consider attenuating the first formant completely since it contains 'important speech information'.

With natural speech, enhancement of speech intelligibility in conditions of masking, such as high noise levels, has been obtained by a 'removal' of the first formant and subsequent infinite peak clipping of the speech signal (Thomas and Niederjohn, 1968; 1970; Thomas and Ravindran, 1971). These studies involved the filtering of PB words with a high pass filter having a cut-off frequency of 1100 Hz and attenuation slopes of 24 dB/octave (Thomas and Niederjohn, 1968) and 12 dB/octave (Thomas and Niederjohn, 1970; Thomas and Ravindran, 1971). The filter characteristics and the stimulus material indicate that the first formant was not 'eliminated', but moderately attenuated.

In view of the improvements in discrimination given by the filtered clipped speech in a masking situation, Thomas and Sparks (1970; 1971) used the same method to test subjects with hearing losses. The results indicated that higher intelligibility scores were obtained for the modified speech and the authors concluded that the use of filtered clipped speech with the first formant 'eliminated' was an effective means of increasing speech discrimination for subjects with various types of hearing impairment. In subsequent research it was discovered that high-pass filtering and clipping offered little advantage over simple high-pass filtering with respect to improvement in intelligibility in noisy conditions (Thomas and Ohley; 1972) and by hearing-impaired subjects (Pfannebecker, 1972).

In summary, studies with synthetic and natural speech indicate that the high sound pressure levels which represent the listening conditions of deaf listeners lead to a spread of masking of high frequency information by low frequencies. One approach to this problem has been the attenuation and the possible elimination of the first formant. It seems that the information content of the first formant of natural speech should be assessed before such measures are taken.

RATIONALE FOR THE STUDY

Research with synthetic speech has indicated that the second formant transition is the predominant carrier of place information for most consonants, and that the first formant transition contains both manner and voicing information. Research with synthetic speech has also indicated that at high sound pressure levels the first formant frequencies produce a masking of second formant transitions.

Research with natural speech has indirectly shown that frequencies in the second formant range are the main contributors to intelligibility, and that first formant frequencies aid in recognition. Place cues have been found in both invariant and vocalic parts of the spectrum, the relative importance of each part depending on contextual factors such as vowel and position. Manner and voicing cues have not been as extensively studied as place cues, but have been associated with the invariant part of the spectrum and with durational features. High levels of presentation have also been found to cause a decrease in intelligibility with the attenuation and 'possible elimination' of first formants alleviating the problem.

There have been as yet no studies which have assessed the contribution of isolated first and second formants to consonant intelligibility and to the perception of the features of place, manner and voicing, and which have investigated the effect of both formants together at comfortable and high sound pressure levels on intelligibility and on the perception of the three phonetic features.

It is the purpose of the present investigation to study the intelligibility and the perception of the features of place, manner and voicing in the consonants /p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, θ, ð, m, n/

and /l/ in filtered CV and VC syllables containing both formants at an overall comfortable level (Condition F12L), the second formant alone (Condition F2), the first formant alone (Condition F1) and both formants at a high overall sound pressure level (Condition F12H). The front vowel /I/ and back vowel /u/ were used in the study, and were mainly chosen because they enabled adequate attenuation of unwanted formants by band-pass filtering. In addition these two vowels differed in the frequency proximity of their first two formants.

It was hypothesized that, of the filter conditions, Condition F12L would be the most intelligible condition, Condition F1 the least intelligible, and that Condition F12H would show a decrease in intelligibility relative to Condition F12L.

With respect to the place feature, it was hypothesized that Conditions F12L and F2 would show the smallest number of place errors, with Condition F1 having the highest number of place errors. It was also hypothesized that Condition F12H would show an increase in place errors relative to F12L and that this increase would be greater for consonants in the /u/ environment than in the /I/ environment because of the proximity of the first two formants in the vowel /u/. It was also hypothesized that, for all conditions containing the second formant, place errors would vary as a function of consonants, vowels and consonant position.

Since manner cues are found in the first formant transition and in the invariant part of the spectrum, it was hypothesized that the smallest number of manner errors would occur in those conditions containing the first formant and invariant energy at the level of the first two formants, namely Conditions F12L and F12H. It was also hypothesized that in Condition F1 manner would be best perceived in those classes of consonants cued by the first formant transition and that in Condition F2, manner would be best perceived in those classes of consonants cued by invariant energy at the level of the second formant.

Voicing cues are found in the voice bar, first formant transition, and in the invariant part of the spectrum. It was hypothesized that the smallest number of voicing errors would occur in those filter conditions containing both of these cues namely Conditions F12L and F12H. It was also hypothesized that in Condition F1, voicing would be best perceived in those consonants cued by low frequencies, and that in Condition F2, voicing would be best perceived in those consonants cued by invariant energy and by durational characteristics at the level of the second formant.

PROCEDURE

This study involved the binaural presentation of eight lists of CV and VC syllables to twenty normal hearing listeners under four filter conditions and one unfiltered condition. The experimental design is summarized in Table 1.

Each list of 68 items was formed by four repetitions of the 17 consonants /p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, θ, ð, m, n, l/ in a syllable with a constant vowel. The eight lists were formed by pairing the consonants pre- and postvocalically with the vowels /I/ and /u/, and by using two speakers to record the stimuli.

The filter conditions are labelled according to the presence or absence of first and second formants:

Condition F12L: Both first and second formants present at an overall comfortable listening level.

Condition F1: First formant present, second formant absent.

Condition F2: Second formant present, first formant absent.

Condition F12H: Both first and second formants present at an overall high sound pressure level.

The terms first and second formant refer to filtered utterances containing predominantly first and/or second formants, in which unwanted formants have been severely attenuated.

In the listening tests, after each stimulus presentation the subject was required to write down the consonant he heard. After the presentation of the lists under all four filter conditions, the subjects were given the same lists in the unfiltered state.

TABLE 1

Summary of Experimental Design

Design: Repeated measures

Variables:

Speakers:	Two males, JD and RS
Consonants:	17 consonants: 6 plosives /p, b, t, d, k, g/ 8 fricatives /f, v, s, z, ʃ, ʒ, θ, ð/ 2 nasals /m, n/ 1 liquid /l/
Vowels:	/I/ and /u/
Positions:	Initial, final.
Conditions:	Condition F12L: F1 at 66 dB SPL F2 at 46 and 50 dB SPL Condition F1: F1 at 66 dB SPL Condition F2: F2 at 46 and 50 dB SPL Condition F12H: F1 at 100 dB SPL F2 at 80 and 84 dB SPL Unfiltered Condition: 66 dB SPL
Subjects:	20 normal-hearing listeners

Composition of the eight lists of 68 items in which each syllable was repeated four times in random order:

List	Speaker	Vowel	Consonant Position
1	JD	/I/	CV
2	JD	/I/	VC
3	JD	/u/	CV
4	JD	/u/	VC
5	RS	/I/	CV
6	RS	/I/	VC
7	RS	/u/	CV
8	RS	/u/	VC

Listening tests:

Each list was given under each of the four filtered conditions and the unfiltered condition, totalling 40 tests (eight lists by five conditions) per subject.

The procedure is described in detail under five headings: 1) speech materials, 2) recording of syllables, 3) spectrographic analysis, 4) preparation of filtered stimuli and 5) subjects and test presentation.

Speech Materials

Seventeen consonants comprising six plosives /p, b, t, d, k, g/, eight fricatives /f, v, s, z, ʃ, ʒ, θ, ð/, two nasals /m, n/ and one liquid /l/ were chosen for the study. Initially the liquid /r/ was also included in the test material, but the very low third formant of the invariant part of the spectrum for this consonant (Potter, Kopp and Green, 1947; O'Connor et al., 1957; Fant, 1968) made the filtering-out of unwanted formants impossible.

The criteria for the selection of the vowels were as follows:

a) a spacing-out in frequency of the first three formants which would permit the extraction of a single formant by filters with fixed centre frequency for the duration of the whole syllable,

b) approximately the same amplitude of the second formant relative to the first, but with differences in the frequency separation of the first two formants for the two vowels. These criteria would enable the study of the effect of the upward spread of masking of the second formant by the first, relative to formant frequency proximity (Martin, Pickett and Colten, 1972).

The vowels /I/ and /u/ were found to meet these criteria (Peterson and Barney, 1952), as shown in Table 2.

TABLE 2

Summary of characteristics of vowels /I/ and /u/ (from Peterson and Barney (1952), male voice).

	Frequency (Hz)			Intensity (dB SPL)		
Vowel	F1	F2	F3	F1	F2	F3
/I/	390	1990	2550	65	45	41
/u/	300	870	2240	65	49	25

Recording of the Syllables

Syllables spoken by two English speaking males, JD and RS, were recorded on tape in a soundproof room (IAC Model 1203-A). The stimuli were spoken into a Unidyne III Model SM 56 microphone, placed eight inches in front of the speaker and above the level of the mouth in order to avoid the breath stream. The speakers were trained to read the stimuli in a natural manner and to maintain a single level of loudness, even intonation and the same duration for all syllables. During the recording of the syllables, the recording level was adjusted to produce a VU meter reading between 0 and -3 dB. The material was recorded on two quarter track of $1\frac{1}{2}$ mil mylar tape, at a speed of 3-3/4 ips on a Revox Model 77-A tape recorder.

Each syllable was recorded three times, and the best recording in terms of clarity, duration and intonation was selected for use in the study. Eight tapes designated as Tapes A were made by dubbing each of the seventeen selected syllables four times in random order to make one list of 68 syllables. Each list contained syllables with one vowel in the same consonant position and were spoken by the same speaker (Table 1).

Spectrographic Analysis

Wide band (300 Hz) spectrograms (Kay Elemetrics Sonagraph Model 6061-B) with an 80-8000 Hz frequency range were made of all the original 136 syllables used in the study. Male speakers were chosen because of the more clearly defined patterns of formants obtained with broad-band analysis. Female voices often show a mixture of formant and harmonic structure (Fant, 1968).

In order to check the accuracy of the filtering, the 272 filtered utterances were also analyzed on the spectrograph. When the results from the wide band analysis were not well defined, narrow band (45 Hz) recordings and contour displays were made. Figure 1 shows an example of the wide band spectrogram of the syllables /pI/, /bI/ and /mI/ taken from List 1.

Before making the spectrograms, the spectrograph was calibrated for frequency, intensity and the duration of the signals with a specially prepared tape. Precautions were always taken to prevent overloading and intermodulation effects by carefully regulating the input level to the spectrograph (Fant, 1962b; Lindblom, 1962).

The spectrograms were first analyzed by segmenting them into sections following Fant's (1962a) specificational system based on manner of production. The vowel formants and formant transitions were grouped in one segment, and the invariant features formed another segment. Thin lines were then drawn marking the centre of the first, second and third formant bars as seen in the spectrograms (Ohman, 1966). The average error found by Lindblom (1962) using the Kay Sonograph with a 300 Hz analyzing bandwidth and a frequency scale of 2000 Hz/inch was less than 50 Hz. The following points were then read for each utterance:

From the vocalic segment (vowel steady state and transition):

(i) The fundamental frequency. This was derived by counting the number of periods for the first centiseconds (½ inch) of the steady state of the vowel formant;

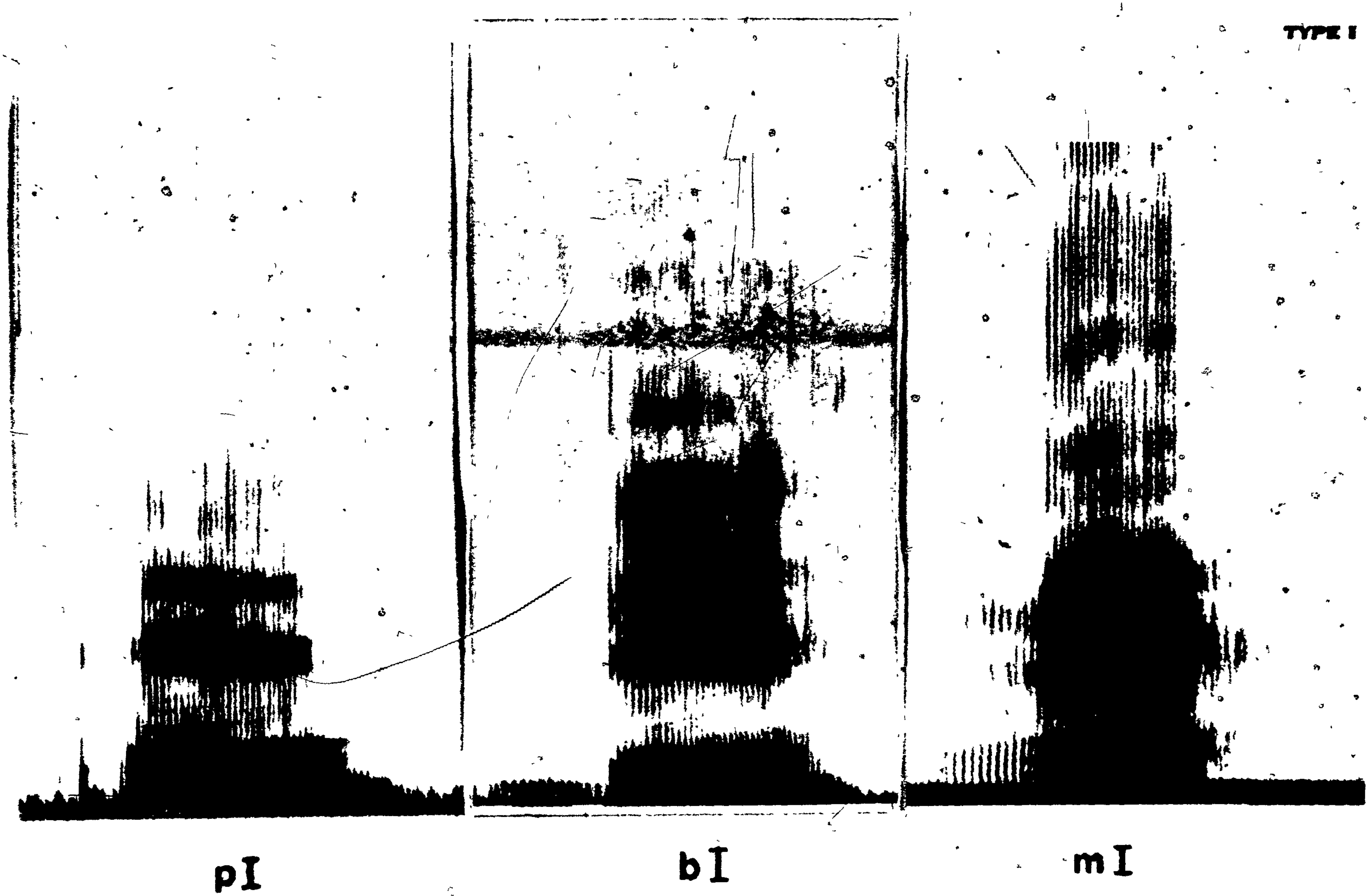

Figure 1: Wide band spectrograms of syllables /pI/, /bI/ and /mI/. (List 1)

(ii) the highest centre frequencies of the first and second vowel formants in their steady or transitional states;

(iii) the lowest centre frequencies of the second and third vowel formants in their steady and transitional states.

From the invariant segment (bursts of stops, friction noises of fricatives, nasal resonances and vocalic like resonances of liquids):

(iv) an attempt was made to read the first, second and third formant frequencies of the invariant part of the energy spectrum, by following Fant's acoustic correlates of manner of production and place of articulation (Fant, 1960; 1962a; 1968). The emphasis was placed on reading the frequencies for the second formant or 'hub' (Potter, Kopp and Kopp, 1966) in order to include these frequencies in the second formant band-passes.

Not all features listed above were well defined for each utterance. Since sounds articulated at the same place show formant pattern transitions that are almost identical (Fant, 1960), the syllables of each list which contained consonants that were produced in the same manner were grouped and the same cut-off frequencies of the filters were used for all syllables in the same group. The nasals were grouped with the plosives as they are often thought of as being nasal stops (Malécot, 1956). The syllables of each list fell into eight groups: /p, b, m/, /t, d, n/, /k, g/, /f, v/, /s, z/, /ʃ, ʒ/, /θ, ð/ and /l/ which formed a separate group. When the second formant of the invariant part could not be seen in any of the consonants of the same group, the reported values (Potter, Kopp and Kopp, 1966) were included. The results of the spectrographic analysis for the 136 syllables are shown in Appendix A.

Preparation of Filtered Stimuli

The preparation of the filtered material consisted of filtering the first and second formants of each utterance separately and recording these on two different quarter tracks of a magnetic tape. Cut-off frequencies of the band-pass filter were determined from the spectrographic analysis of each utterance and the rejection slope of the filter.

Determination of the band-pass filter cut-off frequencies

For the extraction of the second formant, the following frequencies were read from the spectrograms (vocalic and invariant segments) of the syllables of the same group:

a) the lowest and highest second formant centre frequencies,

b) the highest first formant centre frequency,

c) the lowest third formant centre frequency.

An example is shown in Table 3. These frequencies were then plotted alongside the frequency response of the filter (72 dB/octave), as shown in Figure 2.

In a similar manner, for the first formant condition the following frequencies were plotted alongside the frequency response of the filter:

a) the lowest fundamental frequency,

b) the highest first formant centre frequency,

c) the lowest second formant centre frequency (Figure 2).

Centre frequencies were judged adequate because of the relatively narrow bandwidths of the first two formants (Dunn, 1961; Fant, 1967b). Two criteria were set up for the selection of each band-pass. The first concerned the attenuation of unwanted signals. A minimum attenuation of

TABLE 3

Determination of filter cut-off frequencies for first and second formant band-pass (syllables pI-bI-mI, List 1).

Consonant	High F1 (Hz)	Invariant F2 (Hz)	Low F2 (Hz)	High F2 (Hz)	Low F3 (Hz)	Original Band-Pass Cut-Off Frequencies (Hz)	Final Band-Pass Cut-Off Frequencies (Hz)*
p	520	1020 R	2000	2040	2760	F1: 140-560	F1: 140-520
b	560	1020 R	1800	2000	2640	F2: 870-2040	F2: 900-1950
m	560	870 R	1800	2000	2500		

* meeting criteria of minimum attenuation.

R: Reported value (Potter, Kopp and Kopp, 1966)

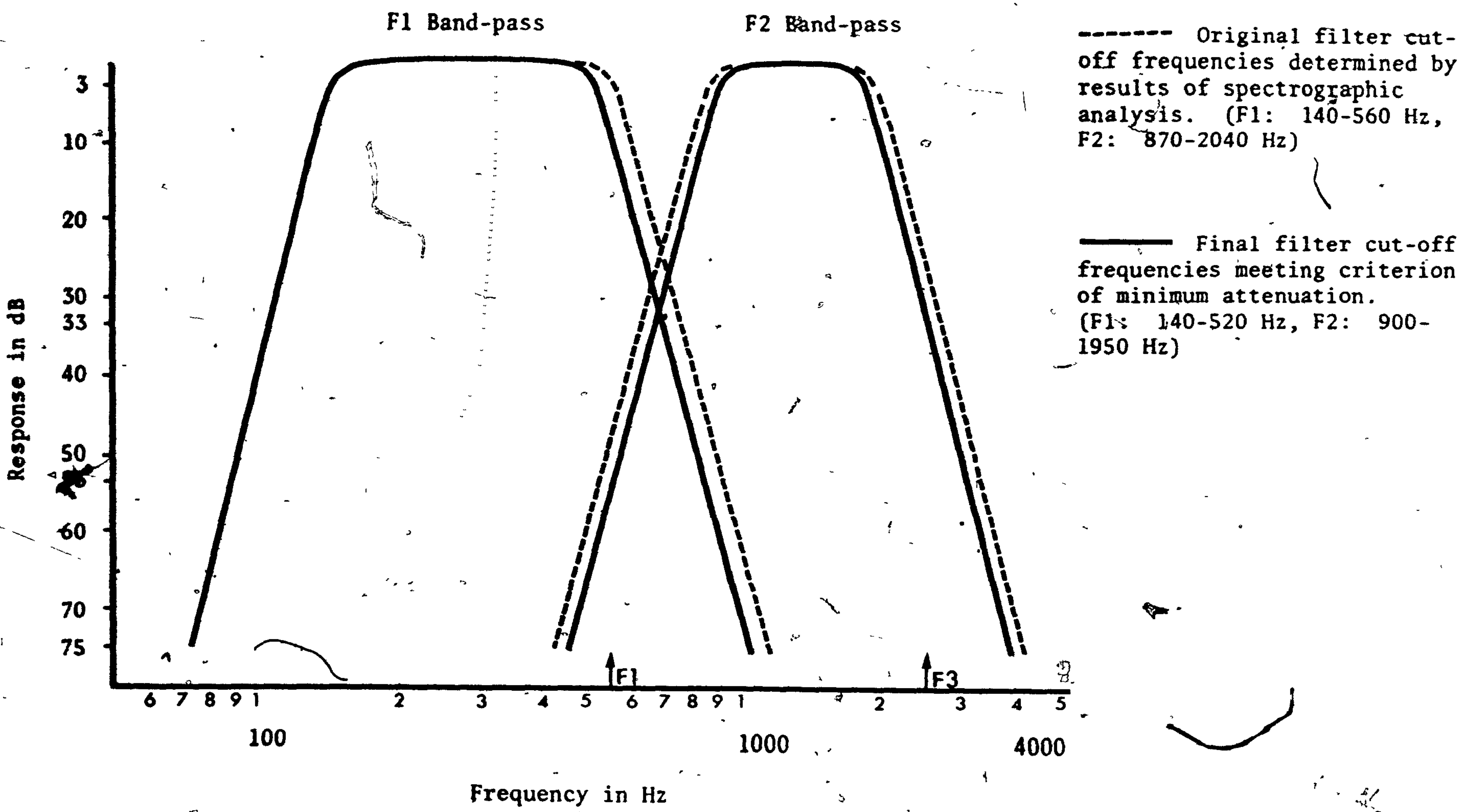

Figure 2: Determination from frequency response of filter (72dB/octave) of cut-off frequencies for first and second formant band-passes for syllable group /pI, bI, mI/. (List 1)

30 dB of third formant centre frequencies and of 50 dB of first formant centre frequencies was required relative to the centre frequency of the second formant band-pass. More attenuation was required of first than of third formants, since first vowel formants are more intense than any other vowel formants (Peterson and Barney, 1952; Fant, 1967b). A minimum attenuation of 50 dB for second formant centre frequencies relative to the centre frequency of the first formant band-pass was also required. When these criteria were met the spectrograms of the filtered utterances clearly displayed only the first or second formants (Figures 3 and 4). In extreme cases, if the previously selected cut-off points did not provide the required attenuation, the cut-off points were shifted to meet the criterion which, of course, resulted in the decrement of the bandwidth of the filter and an attenuation of the previously selected cut-off frequencies (Figure 2).

The second criterion required that the bandwidth for any band-pass should not be less than 200 Hz. This assured reasonable temporal resolution of the filter.

The frequency cut-off points of the first and second formant band-passes are shown in Appendix A and the attenuation provided by the band-passes is given in Appendices B and C.

Preparation of the tapes for the filtered stimuli

The filtering was accomplished by playing a tape on a Revox Model 77-A recorder, whose output was passed through two Rockland active filters Models 1042-F and 1000 connected in series. The filters were Butterworth prototypes and provided a combined roll-off of 72 dB/octave. The output of the filters was recorded on an Ampex Model 620 tape recorder. The overall performance of the filters was first checked on a wave

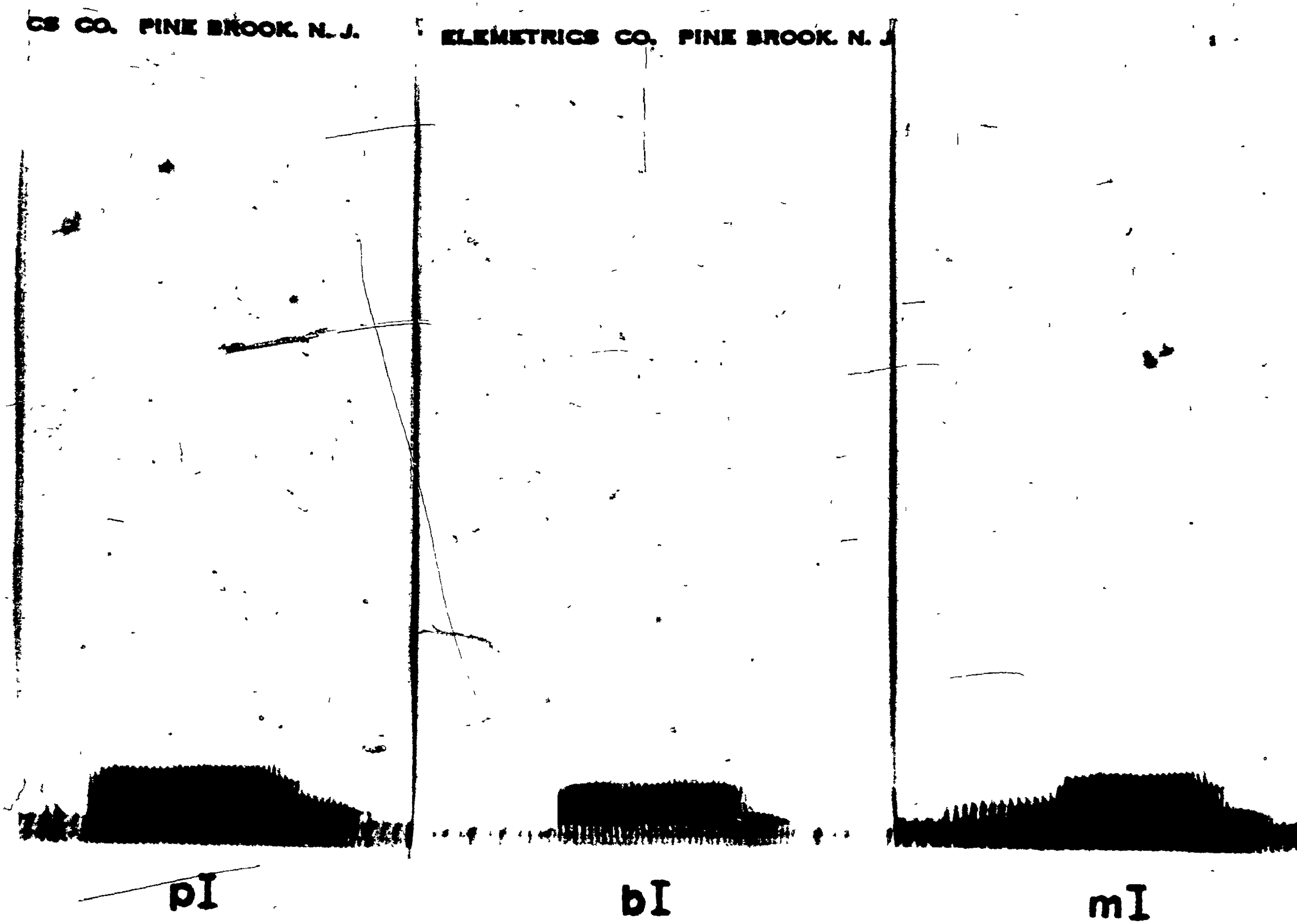

Figure 3: Filtered first formants of syllables /pI/, /bI/ and /mI/. (List 1)

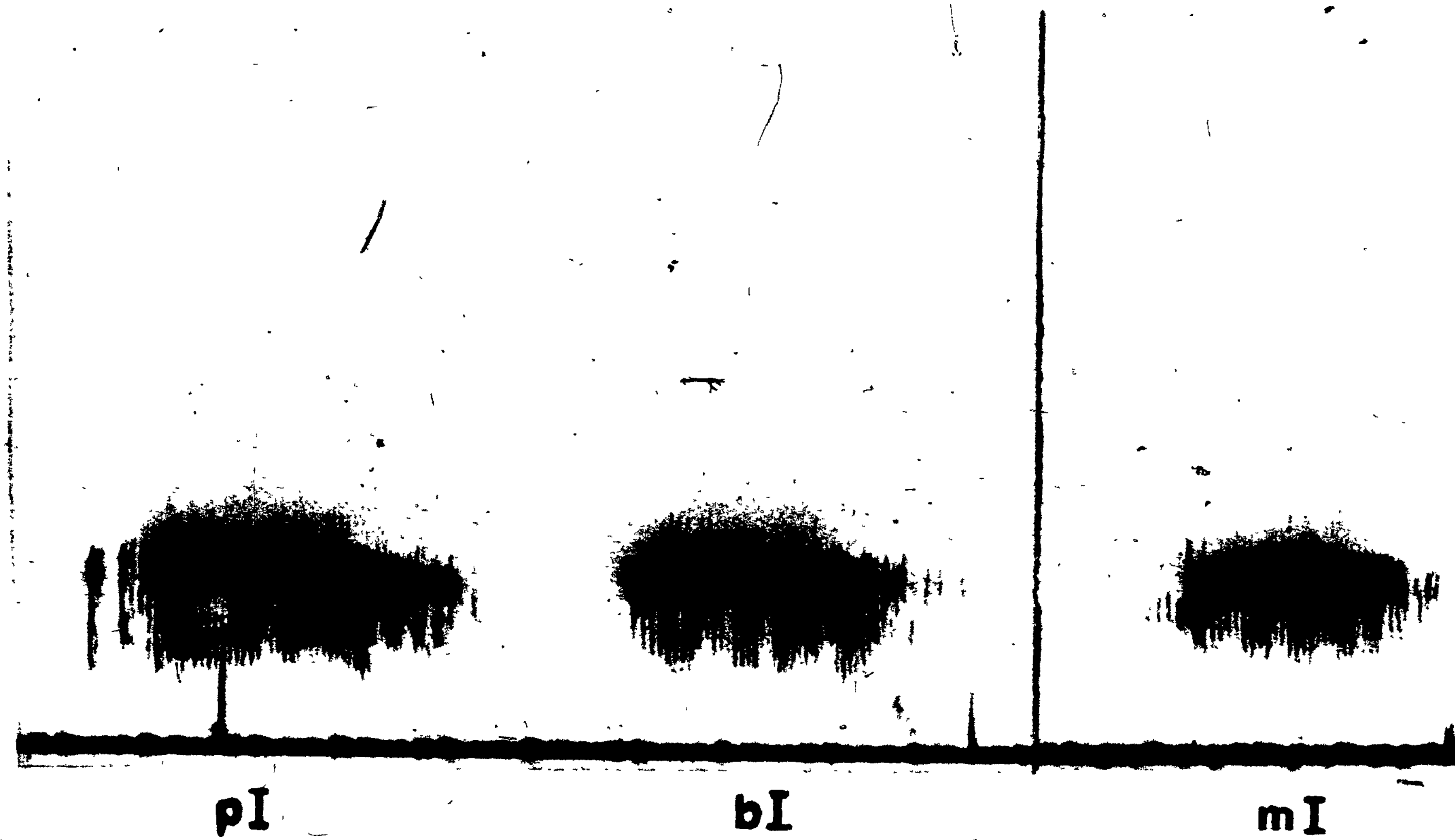

Figure 4: Filtered second formants of syllables /pI/, /bI/ and /mI/. (List 1)

analyzer (General Radio Model 1900A).

Since the final test tape had the first and second formants of each syllable on separate tracks, the filtering process was carried out in two stages. In the first stage, Tapes A containing the eight lists were used as the input to the filters. Second formants were extracted and recorded on one track of a tape while on the second track the unfiltered syllables were dubbed. This tape is referred to as Tape B. Tape B was used in the second stage, which consisted of filtering and recording the first formant on one track of a tape while dubbing the second formant on the other, giving rise to Tape C. This procedure was necessary to ensure synchronization of the two formants. Tape C then contained the first formants on one track and the second formants on the other. A low input level and a considerable weakening of the signal after filtering necessitated the use of amplification by setting the combined gain of the filters to 80 dB for second formants and to 60 dB for first formants. The recording level of the tape recorder was always adjusted to keep VU meter readings between 0 and -3 dB for dubbing and recording of the filtered signals. Four copies of Tape C were made, each containing all filtered lists with an interstimulus duration of three seconds between utterances. These copies were called Tapes D1, D2, D3, D4 and each was used for an equal number of times in the testing of subjects. This ensured that the quality of the recordings would remain acceptable and not deteriorate because of tape stretching. A summary of the preparation of the filtered speech tapes is shown in Table 4. The syllables in Tapes D were not numbered, but each list was divided into four sections containing 17 items. A copy of Tape A was used to present the unfiltered material to the subjects.

TABLE 4

Summary of the preparation of the filtered speech tapes.

Original recording	
↓ Tapes A:	Each syllable dubbed four times in random order.
↓ Tape B:	Second formant on one track, unfiltered signal on other track.
↓ Tape C:	Second formant on one track, first formant on other track.
↓ Tapes D1-4:	Three second interstimulus duration, four sections per list.

Two calibration tones were put on the Tapes D, each matched to give the same VU readings on the tape recorder as the average peak values of the filtered syllables. The calibration tone for the first formant was a 400 Hz pure tone and that for the second formant a 1500 Hz pure tone. These represent the average centre frequencies for the first and second formant band-passes. For the unfiltered material, a 1000 Hz pure tone was used.

Training tape

A training tape was made which included 80 syllables, ten syllables chosen randomly from each filtered list. The training tape thus comprised the filtered first and second formants on separate tracks.

Subjects and Test Presentation

Subjects

The subjects were 20 paid university students, fifteen women and five men, ranging in age from 18 to 25 years. All subjects had normal hearing as indicated by a screening test at 15 dB HL (ANSI-1969), for octave frequencies between 250 and 4000 Hz.

All subjects listened to the syllables binaurally and recorded their responses on a printed response sheet. Each subject was tested for five sessions lasting one hour each. The first session included an additional training period which lasted approximately 25 minutes.

Equipment for listening tests

The taped lists of the filtered syllables were played on a

Revox Model 77-A two channel tape recorder, the two outputs of which were fed into a Grason-Stadler two channel speech audiometer Model 162. Depending upon the test condition, the speech audiometer delivered each signal individually or both signals mixed to the two earphones (Telephonics Model TDH-39 mounted in MX 41/AR cushions), located inside a soundproof room where the subjects listened. It was possible to control the syllable level for first and second formants separately by means of the gain control and attenuators on the speech audiometer.

Instructions for the subjects

Instructions for the listening tests were given by means of a written statement. Subjects were asked to listen carefully to each sound presentation and to write down on a response sheet the consonant that they heard. They were further instructed to make a guess if they were not certain what consonant they had heard, and were urged that under no circumstances should a response space on the sheet be left blank. A card with the 17 consonants used in the study was available for reference throughout the tests. Those who had no experience in phonetics were instructed to write down special characters corresponding to certain sounds, for example /sh/ for /ʃ/, /th/ for /θ/, /Th/, for /ð/ and /j/ for /ʒ/. Examples were given and the examiner made sure the subjects understood the task before starting the test.

Training session

A training session was given to the subjects to provide experience in listening to filtered speech. The training tape was presented at 66 dB SPL, first in Condition F2 and secondly in Condition F1. During the training immediate feedback was given. The subjects were informed that the purpose of the training tape was to give them experience in listening to filtered speech and that these results would not be included in the final score. If a subject left blanks or included consonants which were not used in the study, he was given further instruction.

Test sessions

After the training period, the subject was tested for five sessions. Each of the first four consisted of the presentation of the eight lists under the four filter conditions. In the fifth session the unfiltered material was presented in order to study the intelligibility of the consonants and establish a baseline for comparison.

The four experimental conditions were as follows:

Condition F12L: First formants presented at 66 dB SPL (normal conversational speech) and second formants presented at the corresponding relative amplitudes, namely 46 dB SPL for syllables with the vowel /I/ and 50 dB SPL for syllables with the vowel /u/ (Peterson and Barney, 1952);

Condition F1: First formants presented at 66 dB SPL;

Condition F2: Second formants presented at 46 dB SPL, (syllables with /I/) and at 50 dB SPL (syllables with /u/).

Condition F12H: First formants presented at 100 dB SPL and second formants presented at the corresponding relative amplitudes, namely, 80 dB SPL for syllables with the vowel /I/ and 84 dB for syllables with the vowel /u/.

A pilot study had indicated that a possible learning effect occurred when Condition F2 was preceded by Condition F12L. In order to control for this, four orders of presentation were chosen in such a way that each condition occurred once in each position:

Order 1: Conditions F1, F2, F12L, F12H

Order 2: Conditions F12H, F1, F2, F12L

Order 3: Conditions F12L, F12H, F1, F2

Order 4: Conditions F2, F12L, F12H, F1

For a particular order, the eight lists were randomized within each condition. Five subjects were assigned to each order. Each subject was tested with eight lists during one session and the order of conditions was repeated twice. The test presentation for Subject 1 is given as an example:

Session	Condition F12L	Condition F12H	Condition F1	Condition F2
1	List 7 (1st list)	List 2 (2nd)	List 6 (3rd)	List 3 (4th)
	List 6 (5th)	List 3 (6th)	List 7 (7th)	List 4 (8th)
2	List 1	List 5	List 2	List 7
	List 2	List 7	List 4	List 6
3	List 8	List 6	List 5	List 5
	List 3	List 8	List 1	List 1
4	List 4	List 4	List 8	List 2
	List 5	List 1	List 3	List 8

In the fifth and final session the eight unfiltered lists were administered in random order.

Data Analysis

In preparation of the stimuli, two different speakers had been used to ensure generality of the results. Spearman's coefficient of rank correlation (Ferguson, 1971) was calculated between the two speakers for the paired ranks of the seventeen consonants for each experimental condition. The coefficients of rank correlation were sufficiently high for each condition (Rho: 0.809, 0.775, 0.723 and 0.850 for Conditions F12L, F1, F2 and F12H, respectively) to indicate uniformity in the responses to the stimuli spoken by the two speakers. This uniformity can also be seen by the comparison of consonant ranks listed in Table 5. For all further analyses, therefore, the data for the two speakers were summed.

The data were analyzed in two ways:

(1) Separate analyses of variance (NEW 08V Analysis of Variance Program, Biomedical Computer Programs, W. I. Dixon (Ed.), 1970) were calculated in terms of the number of correct consonant identifications and the number of errors with respect to each of the three phonetic features of place, manner and voicing. The number of correct consonant identifications was used as a measure of intelligibility. The three phonetic features were studied independently, although it is recognized that an incorrect response may differ from the stimulus in one, two or three of the features. Place errors were responses which differed in place of articulation from that of the stimulus, regardless of manner of production and voicing. Manner and voicing errors were determined in a similar way. Table 6 lists the six places of articulation, the four manners of production and the voiced-voiceless distinctions for the 17 consonants used in the present study. It should be noted that these features were not equally distributed for all stimuli. With respect to the place feature, the consonants were

TABLE 5

Consonants ranked with respect to intelligibility for each speaker (JD and RS) under each experimental condition, along with Spearman's rank difference correlation for the four conditions.

	Conditions							
	F12L		F1		F2		F12H	
Consonant	JD	RS	JD	RS	JD	RS	JD	RS
p	11	10	5	8	12	13	10	9
b	3	7	1	9	7	12	1	5
t	16	11	6	2	13	8	14	11
d	6	6	2	4	5.5	6.5	3.5	7
k	1	1	4	3	1	1	2	1
g	2	4	9	5	5.5	1.5	3.5	6
f	10	8	11	6	10	6.5	8	10
v	14	13	10	7	11	9	15	13
s	17	15	12	12	16	14	17	17
z	15	14	14	13	14	10	16	14
ʃ	8	2	17	15	8	2	6	2
ʒ	7	5	16	17	4	5	9	4
θ	13	16	15	14	17	16	12	15
ð	12	17	13	16	15	17	13	16
m	5	9	7	11	2	3	7	8
n	4	3	3	1	3	4	5	3
l	9	12	8	10	9	11	11	12
Correlation Coefficient	Rho = 0.809 t = 5.323*		Rho = 0.775 t = 4.743*		Rho = 0.723 t = 4.076*		Rho = 0.850 t = 6.247*	

*$p < .001$

TABLE 6

Place of articulation, manner of production and voicing for the 17 consonants employed in the present study.

Consonant	Place of Articulation	Manner of Production	Voicing
p	bilabial	plosive	voiceless
b	bilabial	plosive	voiced
t	alveolar	plosive	voiceless
d	alveolar	plosive	voiced
k	velar	plosive	voiceless
g	velar	plosive	voiced
f	labiodental	fricative	voiceless
v	labiodental	fricative	voiced
s	alveolar	fricative	voiceless
z	alveolar	fricative	voiced
ʃ	palatal	fricative	voiceless
ʒ	palatal	fricative	voiced
θ	interdental	fricative	voiceless
ð	interdental	fricative	voiced
m	bilabial	nasal	voiced
n	alveolar	nasal	voiced
l	alveolar	liquid	voiced

divided into three bilabials, six alveolars, two velars, two labiodentals, two palatals and two interdentals. With respect to manner of production there were six plosives, eight fricatives, two nasals and one liquid. Voicing divided the sounds into ten voiced and seven voiceless consonants.

A total of twenty analyses of variance were performed. These are listed in Table 7. The first four comprised four filter conditions-by-17 consonants analyses performed separately on each of the dependent variables, namely the number of correct consonant identifications and the number of place, manner and voicing errors. The subsequent sixteen analyses (Analyses 5-20 of Table 7) comprised a 17 consonants-by-two vowels-by-two positions analysis for each dependent variable and each filter condition.

The .01 level of confidence was adopted with significance levels further evaluated by the conservative test recommended by Greenhouse and Geisser (1959). The conservative test involves the use of reduced degrees of freedom since there is a possibility of violating the assumption of homogeneity of covariance and positively biasing the F-test with a repeated measures design. Multiple comparisons of means for significant main effects and interactions were carried out using Tukey's procedure of honestly significant differences (Winer, 1971). In the case of significant interactions, the critical differences between means were calculated using the combined mean square error and the degrees of freedom computed with the Satterthwaite approximation as recommended by Winer (1971). The .01 level of confidence was also adopted for the results of the Tukey procedure of honestly significant difference.

(2) The consonant confusions were analyzed by the generation

TABLE 7

List of analyses of variance performed in the present study.

	Analysis	Independent Variables	Dependent Variables
All Conditions	1	4 conditions-x-17 consonants	consonant identifications
	2	4 conditions-x-17 consonants	place errors
	3	4 conditions-x-17 consonants	manner errors
	4	4 conditions-x-17 consonants	voicing errors
Condition F1	5	2 vowels-x-2 positions-x-17 consonants	consonant identifications
	6	2 vowels-x-2 positions-x-17 consonants	place errors
	7	2 vowels-x-2 positions-x-17 consonants	manner errors
	8	2 vowels-x-2 positions-x-17 consonants	voicing errors
Condition F2	9	2 vowels-x-2 positions-x-17 consonants	consonant identifications
	10	2 vowels-x-2 positions-x-17 consonants	place errors
	11	2 vowels-x-2 positions-x-17 consonants	manner errors
	12	2 vowels-x-2 positions-x-17 consonants	voicing errors
Condition F12L	13	2 vowels-x-2 positions-x-17 consonants	consonant identifications
	14	2 vowels-x-2 positions-x-17 consonants	place errors
	15	2 vowels-x-2 positions-x-17 consonants	manner errors
	16	2 vowels-x-2 positions-x-17 consonants	voicing errors
Condition F12H	17	2 vowels-x-2 positions-x-17 consonants	consonant identifications
	18	2 vowels-x-2 positions-x-17 consonants	place errors
	19	2 vowels-x-2 positions-x-17 consonants	manner errors
	20	2 vowels-x-2 positions-x-17 consonants	voicing errors

of confusion matrices for each filter condition. The data for the two vowels and two positions were pooled to provide a more stable estimate of the consonant confusions under each filter condition. A spatial representation of the perceptual structure of the 17 consonants was then carried out for each filter condition by using Kruskal's (1964a) non-metric multidimensional scaling procedure (MS-SCAL computer program (Kruskal, 1964b)).

RESULTS

The purpose of this investigation was to determine the intelligibility of 17 consonants presented in initial or final position in the context of the vowel /I/ or /u/, under filter conditions where the subjects heard both formants at a normal listening level (F12L) as compared with conditions containing only the first formant (F1), only the second formant (F2), and both formants at a high level of presentation (F12H). Within each condition the errors in consonant perception were studied as related to the three phonetic features of place of articulation, manner of production and voicing. A fifth condition, in which listeners heard the unfiltered material, enabled a comparison with Condition F12L, which was the filter condition most closely resembling natural speech.

The results for the unfiltered condition are given first. Of the four filter conditions, the results for Condition F12L are always presented first within each analysis, and the results of the remaining three filter conditions are then compared with those for Condition F12L.

Unfiltered Material

When the test stimuli were presented with no filtering, overall intelligibility was high (93.3%) and there was little variation as a function of vowels (93.18% for /I/ and 93.42% for /u/) or position (93.30% for initial and 93.28% for final). The least intelligible consonants were /θ, ð, f, v, p/ and the main substitutions for these consonants involved place confusions (Table 21, p. 92). Because of the high scores, statistics were not performed on the data for the unfiltered material. The results for Condition F12L will be compared to those for the unfiltered material in separate paragraphs at the end of the Analyses of Variance and Analysis of Consonant Confusions sections.

Analyses of Variance

The twenty analyses of variance performed provide a very detailed account of the effect of filtering, and only those comparisons most relevant to the main hypotheses are presented.

The four filter conditions-by-17 consonants analyses were carried out to assess the effect of filtering and the filtering-by-consonant (F x C) interaction. The 17 consonants-by-two vowels-by-two positions analyses were carried out to assess the consonant main effect for each condition, which will be interpreted in the context of significant condition-by-consonant (F x C) interactions in the preceding analyses, and significant vowel-by-consonant (V x C) and significant position-by-consonant (P x C) interactions for the four conditions. Significant vowel-by-position (V x P) and vowel-by-position-by-consonant (V x P x C) interactions, although not of primary importance for this study, are reported briefly in Appendix G.

Summaries of the analyses of variance are presented in Appendix D in which the levels of significance are listed for the reduced degrees of freedom. Only those differences which reached the required level of significance ($p < .01$) are reported as different in the text.

Main effect of filtering

In the four analyses which assessed the consonants under the four filter conditions, all main effects for conditions and consonants and all condition-by-consonant interactions were significant. The differential filter effect on the consonant mean scores and the differences between consonants within each filter condition will be reported in the following sections.

The intelligibility and error percentages for each dependent variable and each condition appear in Table 8, and the rank order of consonant mean scores and results of the Tukey procedure for differences among conditions are shown in Appendix E.

Of the four conditions, Condition F12L in which listeners heard both formants at a normal level of presentation gave the highest overall consonant intelligibility, and Condition F1 which contained only the first formant gave the lowest overall intelligibility. The overall intelligibility of consonants was not different in Condition F2 in which listeners heard only the second formant, from Condition F12H in which both formants were presented at a high level. With respect to the errors for the three phonetic features, Conditions F12L and F2 gave the smallest number of place errors followed by Condition F12H and finally Condition F1 had the highest number of place errors. The number of manner errors was highest in Condition F1, and did not differ for the three remaining conditions. There were fewer voicing errors in both Conditions F12L and F12H than in Conditions F1 and F2 which did not differ from one another with respect to this feature. A comparison of Condition F12L and F2 indicates that the removal of the first formant had the effect of decreasing overall intelligibility and increasing the number of voicing errors, but did not affect the overall perception of manner and place. The removal of the second formant as seen by comparing Conditions F12L and F1, decreased intelligibility and increased the number of errors for all three phonetic

TABLE 8

Intelligibility and error percentages for the four filter conditions.

	Conditions			
	F12L	F1	F2	F12H
Correct consonant identifications	61.7	23.2	56.0	56.4
Place errors	33.9	69.2	35.0	38.8
Manner errors	9.6	35.3	10.8	13.3
Voicing errors	3.9	9.7	11.2	3.8

features. A comparison of Conditions F1 and F2 indicates that both first and second formants contained equal voicing information, but that Condition F2 contained more place and manner information. The effect of a high level of presentation as seen by comparing Conditions F12L and F12H had the effect of decreasing intelligibility and increasing the number of place errors, but did not change the overall perception of manner and voicing. Although the three phonetic features could not easily be compared statistically, it should be noted that the number of place errors exceeded the number of manner errors and voicing errors for all conditions, as might be predicted in terms of the different values for each feature (Table 6, p.61).

Differential filter effects

The percentages of intelligibility and errors for each consonant under the four filter conditions are shown in Table 9. The consonant mean scores across the four conditions were compared in order to investigate which consonants were affected by the removal of the first formant as seen in Conditions F12L and F2, by the removal of the second formant as seen in Conditions F12L and F1, and by amplification as seen in Conditions F12H and F12L. Furthermore, the relative information content of first and second formants with reference to the three phonetic features was assessed for each consonant by comparing Conditions F1 and F2. Multiple comparisons of consonant ~~mean~~ scores across filter conditions were performed with the Tukey procedure. The results appear in Table 9 and in greater detail in Appendix F. Only the comparisons relevant to the hypothesis will be reported in the text. The remaining comparisons are presented in table form in Appendix F, for those readers who may wish to make a more thorough study of the differential effect of filtering on the consonants.

TABLE 9

Intelligibility and error percentages for the three phonetic features, 17 consonants and four filter conditions. (Tukey range: 10.97% for correct consonant identifications, 10.81% for place errors, 8.84% for manner errors, 6.28% for voicing errors.)

	Correct Consonant Identifications				Place Errors				Manner Errors				Voicing Errors			
Consonant	F12L	F1	F2	F12H	F12L	F1	F2	F12H	F12L	F1	F2	F12H	F12L	F1	F2	F12H
p	55.8	28.3	43.4	58.8	42.3	69.2	52.2	40.2	1.7	16.6	4.1	1.4	4.8	13.9	9.1	2.8
b	85.0	40.8	54.7	83.3	13.1	56.1	27.2	15.2	3.8	14.2	23.6	3.8	1.3	9.2	19.1	2.3
t	42.3	33.3	47.8	42.3	55.8	58.1	49.2	57.3	2.7	22.7	6.1	1.4	3.8	12.2	8.8	1.6
d	83.6	42.0	66.4	78.4	16.3	57.5	20.0	20.8	0.8	2.2	13.8	0.0	0.8	1.6	11.7	0.9
k	96.6	36.1	93.3	92.8	3.1	62.8	4.8	7.0	0.2	9.4	0.8	0.0	1.3	7.8	2.3	0.2
g	88.4	27.0	51.7	79.7	9.8	72.0	25.5	17.8	0.9	6.3	7.7	3.6	2.0	2.5	28.4	2.7
f	63.4	24.8	57.0	58.6	32.3	70.6	35.3	37.8	6.3	50.0	10.3	13.1	7.8	17.5	18.4	9.8
v	40.6	24.1	53.3	34.1	58.6	75.6	40.6	65.2	26.7	50.5	8.[illegible]	39.1	5.3	6.9	13.8	6.3
s	23.0	16.9	30.0	13.6	74.8	62.0	65.3	84.4	2.0	39.2	4.4	2.5	5.6	22.5	7.3	5.5
z	37.7	8.9	41.1	22.5	53.3	59.5	53.1	59.1	10.3	54.7	2.2	21.9	6.6	4.4	12.0	5.5
ʃ	77.2	4.7	78.4	78.4	18.6	94.5	11.6	20.8	0.8	32.3	1.3	0.5	6.1	17.7	12.2	2.7
ʒ	74.5	5.6	72.3	69.5	24.8	94.2	12.2	29.8	11.4	54.8	5.0	18.6	1.7	9.4	18.0	3.1
θ	31.6	7.7	25.6	31.3	65.5	92.0	70.9	67.2	6.3	47.8	8.3	5.0	8.4	12.7	13.1	5.3
ð	30.2	8.9	24.7	27.2	65.3	87.7	70.2	67.3	22.0	50.6	14.5	29.2	6.1	7.3	7.5	7.0
m	75.2	22.5	83.3	63.1	10.8	69.1	12.7	19.2	16.6	42.2	9.2	23.6	0.9	7.2	0.3	1.3
n	86.6	40.3	75.5	77.7	6.6	46.9	11.9	15.8	7.3	28.9	16.7	9.5	0.2	4.2	2.0	1.1
l	56.7	22.5	52.7	47.2	25.0	47.7	32.0	35.5	43.3	77.5	47.5	52.8	3.8	8.8	5.9	7.3

A comparison of Conditions F12L and F2 indicates that the plosives /p, b, d, g/ and the nasal /n/ were less intelligible when only the second formant was present, while the voiced fricative /v/ was more intelligible with only the second formant. The remaining consonants showed no change in intelligibility when the first formant was removed. The absence of the first formant decreased the number of place errors in /v/ and /ʒ/, but increased the number of place errors for the voiced plosives /b/ and /g/. Manner errors also decreased in Condition F2 relative to F12L for /v/, but increased for the voiced plosives /b/ and /d/, and the nasal /n/. With respect to voicing errors, the absence of the first formant had the effect of increasing the number of errors for /b, d, g, f, v, ʒ/. These comparisons indicate that the first formant contained some place information for /b/ and /g/, manner information for /b, d, n/ and voicing information for the voiced plosives and the fricatives /f, v, ʒ/. The presence of the first formant was detrimental to the perception of place and manner for /v/ and of place for /ʒ/.

A comparison of Conditions F12L and F1 indicates that all consonants except /t/ and /s/ became less intelligible in the absence of the second formant. With respect to the three phonetic features, the removal of the second formant increased the number of place errors for all consonants except /t/ and /z/, the number of manner errors for all consonants except /d/ and /g/, and the number of voicing errors for the voiceless plosives, the voiceless fricatives /f, s, ʃ/ and for the voiced consonants /b, ʒ, m/. These comparisons show that the second formant contributed place and manner information for almost all of the consonants and voicing information for about half of the 17 consonants.

A comparison of Conditions F1 and F2 indicates that all consonants were more intelligible in the condition containing the second formant alone than in the condition containing the first formant alone. In addition, place errors were fewest in Condition F2 for all consonants except /t, s, z/, which showed no difference in place errors between F1 and F2. There were fewer manner errors in F2 for the voiceless plosives /p, t/, all fricatives, the nasals and /l/, but more manner errors for the voiced plosives /b, d/, and no difference for the plosives /k/ and /g/. There were more voicing errors in Condition F2 than F1 for the voiced plosives /b, d, g/ and the voiced fricatives /v, z, ʒ/, but fewer errors for /s/ and /m/, with no difference for the remaining consonants. For the two formants, therefore, the second contained more place information for the majority of consonants, more manner information for the voiceless plosives /p, t/, all fricatives, the nasals and /l/, and more voicing information only for /s/ and /m/; while the first formant contained more manner information for the voiced plosives /b, d/, and more voicing information for the voiced plosives /b, d, g/ and voiced fricatives /v, z, ʒ/. Both formants contained as much information with respect to place for /t, s, z/, manner for /k, g/ and voicing for /p, t, k, f, ʃ, θ, ð, n, l/.

The effects of relatively high sound pressure level of the two formants can be seen by comparing Conditions F12L and F12H. Only the consonants /z/ and /m/ were less intelligible in Condition F12H than in F12L. With respect to place errors, there were no differences in the number of place errors for individual consonants. Manner errors increased in Condition F12H for /v, z, l/ and the individual consonants showed no differences between F12L and F12H with respect to voicing errors. Increasing sound pressure level had the effect of decreasing intelligibility and increasing manner errors for a minority of consonants, but did not increase the number of place or voicing errors for any of the consonants.

Differences between consonants within each filter condition

The effect of consonants within filter conditions was significant for each of the four dependent variables with the exception of voicing errors for Conditions F12L and F12H. In order to assess the relative information of each condition for each consonant, multiple comparisons were carried out using the Tukey procedure. The detailed results are shown in Appendix E and Table 10 shows the significant consonant groupings for each dependent variable under each condition.

The consonant groups obtained for Condition F12L indicate that the intelligibility of the voiceless plosive /k/, the voiced plosives, the nasals and the fricatives /ʃ, ʒ/ tended to be higher than that of the voiceless plosives /p, t/ and the fricatives /s, ð, θ, z, v/, the intelligibility of /f/ and /l/ being of intermediate accuracy. Consonant intelligibility was influenced by the accuracy with which the place feature was perceived, as seen by the similar consonant groupings for place errors and intelligibility. With respect to the manner feature, there were relatively fewer errors for all plosives, the voiceless fricatives and the nasal /n/ than for the nasal /m/, the voiced fricatives and the liquid /l/. It should be noted that any error response for the liquid /l/ was considered a manner error since /l/ was the only consonant of its class. The consonant /k/ was the most intelligible consonant in Condition F12L with the least number of place and manner errors. No significant differences among consonants occurred with respect to the voicing feature in F12L.

TABLE 10

Groups of consonants not differing significantly from one another as found with the Tukey procedure.

CONDITION F12L		CONDITION F1
high k g n b d ʃ m ʒ f l p t v z θ ð s low	Correct Consonant Identifications	high d b n k t p g f v l m s z ð θ ʒ ʃ low
low k n g m b d ʃ ʒ l f p z t v ð θ s high	Place Errors	low n l b d t z s k m p f g v ð θ ʒ ʃ high
k ʃ d g p s t b θ f n z ʒ m ð v l	Manner Errors	d g k b p t n ʃ s m θ f v ð z ʒ l
No significant differences among consonants.	Voicing Errors	d g n z v m ð k l b ʒ t θ p f ʃ s

CONDITION F2		CONDITION F12H
high k m ʃ n ʒ d f b v l g t p z s θ ð low	Correct Consonant Identifications	high k b g ʃ d n ʒ m p f l t v θ ð z s low
low k ʃ n ʒ m d g b l f v t p z s ð θ high	Place Errors	low k b n g m ʃ d ʒ l f p t z v θ ð s high
k ʃ z p s ʒ t g θ v m f d ð n b l	Manner Errors	k d ʃ t p s g b θ n f ʒ z m ð v l
m n k l s ð t p d z ʃ θ v ʒ f b g	Voicing Errors	No significant differences among consonants.

The consonant groups obtained for Condition F1 showed that all plosives and the nasal /n/ formed the most intelligible groups of consonants and the fricatives /ʃ, ʒ, θ, ð, z, s/ the least intelligible group of consonants. The intelligibility of /m, l, v, f/ was of intermediate accuracy. Similar consonant groupings occurred with respect to place errors, with the exception of /l, z, s/, being part of the group with the smallest number of place errors. Manner errors were relatively few for plosives and high for most fricatives and the liquid /l/, the identification of manner for the nasals and for the fricatives /ʃ/ and /s/ being of intermediate accuracy. The gross groupings for voicing indicated that accuracy was highest for all voiced sounds together with the voiceless plosive /k/, and the lowest for the remaining voiceless consonants.

The groups of consonants obtained for Condition F2 showed that the plosive /k/, the nasals, the fricatives /ʃ, ʒ/ and the plosive /d/ were the most intelligible consonants with the fewest place errors, and that the fricatives /s, z, θ, ð/ were the least intelligible consonants with the highest number of place errors. No systematic clustering of plosives or fricatives occurred with respect to intelligibility and place errors. Manner errors were fewest for all fricatives except /ð/, for the voiceless plosives, for /g/ and the nasal /m/, and were highest for /d, ð, n, b, l/. No systematic grouping of voiced and voiceless consonants occurred with respect to voicing errors. Voicing was well identified in the nasals, the liquid /l/, the voiceless plosives and the fricatives /s, ð/. The consonants /ʒ, f, b, g/ gave the most voicing errors in Condition F2.

The groups of consonants in Condition F12H were very similar to those obtained in Condition F12L.

Effects of vowel context

The dependent variables for which the effect of vowels and the vowel-by-consonant interaction were significant appear in Table 11, and the intelligibility and error percentages for each consonant in the two vowel environments are shown in Tables 12 to 15 for the four dependent variables and four filter conditions. In Tables 12 to 15 the significant differences in mean scores between the two vowel environments, as obtained with the Tukey procedure, are indicated for the dependent variables and conditions for which the V x C interaction was significant.

In Condition FI2L significant differences between vowels were found for all four dependent variables, and the vowel-by-consonant interaction was significant for all dependent variables except voicing errors. Overall intelligibility and the overall perception of place, manner and voicing were more acute for consonants in the /I/ environment than in the /u/ environment. With respect to individual consonants, intelligibility was higher and place errors fewer with /I/ for /p, f, ʃ, ʒ/ and with /u/ for /d, s, z, θ/. The consonant /l/ was more intelligible with /I/, but showed no significant difference in the number of place errors between the two vowel environments. Although the overall perception of manner was more acute for the /I/ environment, analysis of individual consonants reveals that this was only the case for /v, ʒ, m, l/, while only /ð/ had fewer manner errors with /u/.

In Condition FI, differences between vowels were found for all dependent variables except voicing errors, and the vowel-by-consonant interaction was significant for manner and voicing errors. Overall intelligibility and overall perception of the place feature were more acute for consonants in the /I/ environment, than in the /u/ environment. With

TABLE 11

Significant V and V x C interactions.

Dependent Variable	F12L V	F12L V x C	F1 V	F1 V x C	F2 V	F2 V x C	F12H V	F12H V x C
	FILTER CONDITIONS							
Correct Consonant Identifications	x	x	x			x		x
Place Errors	x	x	x			x		x
Manner Errors	x	x	x	x		x	x	x
Voicing Errors	x			x			x	

TABLE 12

Intelligibility percentages for each consonant in the two vowel environments.

	F12L		F1		F2		F12H	
	/I/	/u/	/I/	/u/	/I/	/u/	/I/	/u/
p	68.8	42.8*	31.6	25.0	60.9	25.9*	69.1	48.4*
b	91.6	78.4	42.5	39.1	53.8	55.6	84.4	82.8
t	45.0	39.7	29.4	37.2	49.7	45.9	48.1	36.6
d	73.4	93.8*	42.5	41.6	52.5	80.3*	65.0	91.9*
k	95.0	98.1	47.5	24.7	92.5	94.1	87.5	98.1
g	86.9	90.0	42.8	11.3	57.2	46.3	79.4	80.0
f	72.8	54.1*	28.1	21.6	63.4	50.6	75.3	41.9*
v	41.3	40.0	20.9	27.2	51.6	55.0	37.2	30.9
s	15.6	30.3*	19.4	14.4	15.3	44.7*	9.7	17.5
z	27.8	47.5*	9.1	8.8	37.5	44.7	14.1	30.9*
ʃ	97.2	57.2*	7.5	1.9	91.3	65.6*	98.4	58.4*
ʒ	97.8	51.3*	8.1	3.1	84.1	60.6*	92.5	46.6*
θ	20.0	43.1	5.6	9.7	11.9	39.4*	6.9	55.6*
ð	26.5	34.1	8.8	9.1	19.1	30.3	16.9	37.5*
m	81.6	68.8	25.3	19.7	78.4	88.1	68.1	58.1
n	81.6	91.6	50.9	29.7	66.6	84.4*	68.4	86.9*
l	69.1	44.4*	22.8	22.2	68.4	36.9*	60.6	33.8*

*, significant difference between the two vowel environments ($p < .01$).

Tukey range across vowels: 13.5% for F12L, 13.63% for F2, 14.0% for F12H.

TABLE 13

Place error percentages for each consonant in the two vowel environments.

	F12L		F1		F2		F12H	
	/I/	/u/	/I/	/u/	/I/	/u/	/I/	/u/
p	29.7	55.0*	65.0	73.4	32.8	71.6*	30.0	50.3*
b	7.5	18.8	53.8	58.4	20.3	34.1*	12.8	17.5
t	53.1	58.4	60.9	55.3	46.6	51.9	51.9	62.8
d	26.6	5.9*	56.9	58.1	28.8	11.3*	35.0	6.6*
k	5.0	1.3	51.6	74.1	5.9	3.8	12.2	1.9
g	13.1	6.6	56.6	87.5	19.4	31.6	20.3	15.3
f	25.6	39.1*	66.9	74.4	31.9	38.8	23.4	52.2*
v	57.5	59.7	78.8	72.5	44.7	36.6	61.3	69.1
s	84.1	65.6*	64.7	59.4	83.8	46.9*	90.0	78.8
z	60.6	45.9*	56.3	62.8	57.8	48.4	63.4	54.7
ʃ	2.5	34.7*	92.2	96.9	1.3	21.9*	1.3	40.3*
ʒ	1.6	48.1*	91.6	96.9	0.3	24.1*	6.6	53.1*
θ	79.1	51.9*	94.1	90.0	86.3	55.6*	92.2	42.2*
ð	69.4	61.3	89.4	85.9	75.9	64.4	78.8	55.9*
m	9.7	11.9	72.2	65.9	21.3	4.1*	25.9	12.5
n	11.6	1.6	42.2	51.6	18.1	5.6	24.1	7.5*
l	19.4	30.6	50.3	45.0	25.9	38.1	32.8	38.1

* significant difference between the two vowel environments ($p < .01$).

Tukey range across vowels: 12.75% for F12L, 13.13% for F2, 13.75% for F12H.

TABLE 14

Manner error percentages for each consonant in the two vowel environments.

	F12L		F1		F2		F12H	
	/I/	/u/	/I/	/u/	/I/	/u/	/I/	/u/
p	1.3	2.2	15.0	18.1	6.6	1.6	0.9	1.9
b	0.9	6.6	14.4	14.1	25.0	22.2	1.6	5.9
t	2.5	2.8	20.0	25.3	8.8	3.4	0.6	2.2
d	0.9	0.6	1.9	2.5	17.5	10.0	0.0	0.0
k	0.3	0.0	5.0	13.8	1.3	0.3	0.0	0.0
g	0.9	0.9	3.8	8.8	7.8	7.5	6.9	0.3
f	3.4	9.1	46.9	53.1	9.1	11.6	3.8	22.5*
v	6.6	46.9*	53.1	47.8	4.7	12.8	24.7	53.4*
s	0.3	3.8	29.7	48.8*	4.4	4.4	0.9	4.1
z	12.8	7.8	55.9	53.4	0.9	3.4	27.2	16.6*
ʃ	0.6	0.9	29.4	35.3	0.6	1.9	0.3	0.6
ʒ	0.0	22.8*	48.1	61.6*	0.3	9.7*	1.6	35.6*
θ	7.2	5.3	50.0	45.6	13.4	3.1*	4.4	5.6
ð	31.6	12.5*	60.6	40.6*	21.3	7.[illegible]	36.6	21.9*
m	10.0	23.1*	26.3	58.1*	8.4	10.0	12.2	35.0*
n	6.9	7.8	14.1	43.8*	18.4	15.0	8.8	10.3
l	30.9	55.6*	77.2	77.8	31.6	63.1*	39.4	66.3*

* significant difference between the two vowel environments ($p < .01$)

Tukey range across vowels: 7.25% for F12L, 10.5% for F1, 8.38% for F2, 9.25% for F12H.

TABLE 15

Voicing error percentages for each consonant in the two vowel environments.

	F12L		F1		F2		F12H	
	/I/	/u/	/I/	/u/	/I/	/u/	/I/	/u/
p	3.1	6.6	18.8	9.1*	10.0	8.1	3.4	2.2
b	1.6	0.9	13.1	5.3*	26.6	11.6	3.4	1.3
t	1.9	5.6	12.2	12.2	9.4	8.1	0.9	2.2
d	0.9	0.6	0.3	2.8	15.6	7.8	0.3	1.6
k	1.3	1.3	6.6	9.1	2.5	2.2	0.3	0.0
g	0.3	3.8	1.9	3.1	29.7	27.2	0.3	5.0
f	3.8	11.9	18.8	16.3	14.1	22.8	4.4	15.3
v	5.6	5.0	10.3	3.4*	10.9	16.6	8.4	4.1
s	3.4	7.8	9.7	35.3*	3.8	10.9	1.3	9.7
z	7.8	5.3	4.4	4.4	10.0	14.1	6.9	4.1
ʃ	0.3	11.9	10.9	24.4*	7.8	16.6	0.3	5.0
ʒ	0.6	2.8	10.9	7.8	15.9	20.0	2.8	3.4
θ	2.2	14.7	5.3	20.0*	12.2	14.1	3.1	7.5
ð	6.3	5.9	8.8	5.9	6.9	8.1	6.6	7.5
m	0.3	1.6	7.2	7.2	0.0	0.6	1.3	1.3
n	0.0	0.3	2.8	5.6	2.5	1.6	0.6	1.6
l	0.3	7.2	4.7	12.8*	3.1	8.8	0.3	14.4

* significant difference between the two vowel environments ($p < .01$).

Tukey range across vowels: 6.63% for F1.

respect to manner of production, only the consonants /s, ʒ, m, n/ showed fewer errors with /I/, as was the case with /u/ for /ð/. Voicing was better perceived with /I/ for /s, θ, ʃ, l/ and with /u/ for /p, t, v/, thus cancelling out the vowel main effect.

In Condition F2, there were no differences between vowels for any of the dependent variables and the vowel-by-consonant interaction was significant for all dependent variables except voicing errors. With respect to intelligibility and to the perception of the place feature, large differences for the two vowels occurred for some consonants. Intelligibility and the perception of the place feature were significantly more acute in the /I/ environment for /p, ʃ, ʒ/ and in the /u/ environment for /d, s/. In addition, there were fewer place errors with /u/ for /b/ and with /I/ for /m/, but the intelligibility of these two consonants was not different for these two vowels. With respect to manner errors, there were fewer errors with /I/ for /ʒ, l/, the latter consonant also being more intelligible in the /I/ environment, and with /u/ for /θ, ð/. Thus, in Condition F2 intelligibility was highest and the perception of place and manner more acute for some consonants with /I/ and for others with /u/, thus cancelling out any vowel main effect.

In Condition F12H, differences between vowels were found for manner and voicing errors, and the vowel-by-consonant interaction was significant for all dependent variables except voicing errors. The overall perception of manner and voicing was more acute for consonants in the /I/ environment than in the /u/ environment. Intelligibility was higher and place errors fewer with /I/ for /p, f, ʃ, ʒ/ and with /u/ for /d, θ, ð, n/. The consonant /z/ was more intelligible with /u/ and /l/ with /I/. With respect to manner errors, there were fewer errors with /I/ for /f, v, ʒ, m, l/ and with /u/ for /z, ð/.

In summary, for the three conditions which included the second formant, the vowel-by-consonant interactions followed the same trends for consonant intelligibility and for the perception of place. In Condition F1, overall intelligibility and the overall perception of place were most acute in the /I/ environment, and the interactions non-significant. Generally, manner was as well perceived for most consonants in both vowel environments with the exception of a minority of consonants which showed fewer manner errors with /I/, the differences being large enough to make the vowel effect significant in Conditions F12L, F1 and F12H. In Condition F2, each vowel enhanced the perception of manner for an equal number of consonants and the vowel main effect was cancelled out.

Differences in vowel context effects for Conditions F12L and F12H.

The results for the consonant main effects (Table 8) indicated that consonants became less intelligible and the perception of place and manner less acute in Condition F12H relative to F12L. It has been suggested that an upward spread of masking of second formant transitions may occur in the presence of low frequency energy, such as first formant energy, and that the frequency proximity of vowel formants may be a factor in determining the degree of masking (Danaher, Osberger and Pickett, 1973). As the vowels used in this study differed in the relative proximity of their first two formants, /I/ having a larger frequency separation between the two formants than /u/, the scores for consonants exhibiting deteriorations in F12H were examined in both vowel environments to study the effect of the vowels on the differences observed. The percentage differences in intelligibility, and place and manner errors are shown in Table 16 for consonants in both

TABLE 16

Percentage differences between Conditions F12L and F12H for correct consonant identifications and place and manner errors for the 17 consonants in both vowel environments.

	F12L - F12H		F12H - F12L			
	Correct Consonant Identifications		Place Errors		Manner Errors	
	/I/	/u/	/I/	/u/	/I/	/u/
p	- 0.3	- 5.6	0.3	- 4.7	- 0.4	- 0.3
b	7.2	- 3.8	5.3	- 1.3	0.7	- 0.7
t	- 3.1	3.1	- 1.2	4.4	- 1.9	- 0.6
d	8.4	1.9	8.4	0.7	- 0.9	- 0.6
k	0.9	10.6	0.6	7.2	- 0.3	0.0
g	7.5	10.0	7.2	6.6	6.0	- 0.6
f	- 2.5	12.2	- 2.2	13.1	0.4	13.4
v	4.1	9.1	3.8	9.4	6.6	6.5
s	5.9	12.8	5.9	13.2	0.6	0.3
z	13.7	16.1	2.8	8.8	14.4	8.8
ʃ	- 1.2	- 1.2	- 1.2	5.6	- 0.3	- 0.3
ʒ	5.3	4.7	5.0	5.0	1.6	12.8
θ	13.1	-12.5	13.1	- 9.7	2.8	0.3
ð	9.4	- 3.4	9.4	- 5.4	5.0	9.4
m	13.5	10.7	10.7	0.6	2.2	11.9
n	13.2	4.7	12.5	5.9	1.9	2.5
l	8.5	10.6	13.5	7.5	8.5	10.7

vowel environments. Examination of the differences in scores in Table 16 indicates that deteriorations occurred in Condition F12H relative to F12L for consonants in both vowel environments and that, in general, the extent of the differences was similar for the two vowels.

Effects of consonant position

The dependent variables for which the effect of position and the position-by-consonant interaction were significant appear in Table 17. The intelligibility and error percentages for each consonant in the initial and final positions are shown in Tables 18 to 20, for the dependent variables and filter conditions, for which the position effect and position-by-consonant interaction were significant. In Tables 18 to 20, the differences in mean scores between the two consonant positions as obtained with the Tukey procedure are indicated.

In Condition F12L, differences between the two positions were found for all dependent variables except voicing errors. Overall consonant intelligibility and the overall perception of the place feature were most acute for final consonants, and manner best perceived in initial consonants. The position-by-consonant interaction was significant for all dependent variables except voicing errors. There were more correct identifications and fewer place errors for /p, t, d, z/ in the final position and for /v, ð/ in the initial position. The consonant /g/ also showed more correct identifications in the final position, but did not differ in the number of place errors between the two positions. With respect to the perception of manner, there was no difference between the two positions for most consonants with the exception of /v, s, z, θ, ð/ which had fewer manner errors in the initial position, the differences being large enough to make the position

TABLE 17

Significant P and P x C interactions.

Dependent Variable	F12L P	F12L P x C	F1 P	F1 P x C	F2 P	F2 P x C	F12H P	F12H P x C
Correct Consonant Identifications	x	x	x	x	x	x	x	x
Place Errors	x	x	x	x	x	x	x	x
Manner Errors	x	x	x	x	x	x		x
Voicing Errors								

TABLE 18

Intelligibility percentages for each consonant in the two positions.

	F12L		F1		F2		F12H	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final
p	25.0	86.6*	21.6	35.0*	17.8	69.1*	27.2	90.3*
b	86.3	83.8	38.4	43.1	54.7	54.7	77.8	88.8
t	31.9	52.8*	31.6	35.0	37.2	58.4*	30.6	54.1*
d	71.3	95.9*	32.2	51.9*	51.9	80.9*	68.8	88.1*
k	95.3	97.8	58.8	13.4*	90.0	96.6	88.4	97.2
g	83.1	93.8	25.9	28.1	44.7	58.8	65.6	93.8*
f	64.7	62.2	39.7	10.0*	53.3	60.3	56.9	60.3
v	54.7	26.6*	39.7	8.4*	50.9	55.6	46.9	21.3*
s	23.4	22.5	20.6	13.1	30.3	29.7	13.8	13.4
z	20.3	55.0*	10.6	7.2	23.8	58.4*	10.3	34.7*
ʃ	76.9	77.5	4.1	5.3	77.2	79.7	68.4	88.4*
ʒ	71.3	77.8	5.0	6.3	70.9	73.8	56.9	82.2*
θ	35.0	28.1	6.3	9.1	31.6	19.7	34.1	28.4
ð	42.5	17.8*	13.1	4.7	38.4	10.9*	34.4	20.0*
m	68.8	81.6	30.6	14.4*	82.5	84.1	51.9	74.4*
n	80.6	92.5	41.9	38.8	71.9	79.1	69.7	85.6*
l	57.5	55.9	35.0	10.0*	39.7	65.6*	37.2	57.2*

* significant difference between the two positions ($p < .01$).

Tukey range across positions: 13.13% for F12L, 12.75% for F1, 15.88% for F2, 14.38% for F12H.

TABLE 19

Place error percentages for each consonant in the initial (I) and final (F) positions.

	F12L		F1		F2		F12H	
	I	F	I	F	I	F	I	F
p	74.7	10.0*	76.9	61.6*	81.3	23.1*	72.2	8.1*
b	12.2	14.1	55.6	56.6	31.9	22.5	19.1	11.3
t	67.2	44.4*	61.9	54.4	60.9	37.5*	69.4	45.3*
d	28.8	3.8*	67.5	47.5*	35.3	4.7*	31.3	10.3*
k	4.7	1.6	40.0	85.6*	9.7	0.0	11.6	2.5
g	16.9	2.8*	73.8	70.3	50.6	0.3*	34.1	1.6*
f	30.9	33.8	55.0	86.3*	38.8	31.9	37.8	37.8
v	45.3	71.9*	60.0	91.3*	48.4	32.8*	51.9	78.4*
s	75.6	74.1	66.6	57.5	65.0	65.6	85.6	83.1
z	77.2	29.4*	68.8	50.3*	73.8	32.5*	77.2	40.9*
ʃ	22.2	15.0	95.0	94.1	13.8	9.4	31.3	10.3*
ʒ	28.8	20.9	94.7	93.8	15.0	9.4	42.8	16.9*
θ	61.3	69.7	93.1	90.9	63.8	78.1	64.1	70.3
ð	51.3	79.4*	81.6	93.8	52.5	87.8*	57.8	76.0*
m	6.6	15.0	57.8	80.3*	10.9	14.4	15.9	22.5
n	6.9	6.3	51.6	42.2	10.0	13.8	22.2	9.4
l	19.4	30.6	44.1	51.3	39.1	25.0	38.1	32.8

* significant difference between the two positions ($p < .01$).

Tukey range across positions: 12.63% for F12L, 14.0% for F1, 14.75% for F2, 13.75% for F12H.

TABLE 20

Manner error percentages for each consonant in the initial (I) and final (F) positions.

	F12L		F1		F2		F12H	
	I	F	I	F	I	F	I	F
p	3.4	0.0	12.2	20.9	7.8	0.3	2.5	0.3
b	5.3	2.2	27.2	1.3*	39.4	7.8*	7.2	0.3
t	1.9	3.4	19.1	26.3	7.5	4.7	1.6	1.3
d	1.6	0.0	2.2	2.2	23.8	3.8*	0.0	0.0
k	0.0	0.3	14.4	4.4	1.6	0.0	0.0	0.0
g	1.9	0.0	7.5	5.0	15.0	0.3*	7.2	0.0
f	3.8	8.8	27.8	72.2*	17.2	3.4*	15.0	11.3
v	1.9	51.6*	30.3	70.6*	4.4	13.1	12.5	65.6*
s	0.9	3.1	27.5	50.9*	7.5	1.3	1.9	3.1
z	2.8	17.8*	29.7	79.7*	0.9	3.4	15.9	27.8*
ʃ	0.6	0.9	25.6	39.1*	1.6	0.9	0.3	0.6
ʒ	10.0	12.8	43.1	66.6*	6.6	3.4	26.6	10.6*
θ	1.6	10.9*	34.1	61.6*	8.8	7.8	2.5	7.5
ð	6.3	37.8*	30.6	70.6*	3.1	25.9*	20.3	38.1*
m	25.9	7.2*	41.6	42.8	7.2	11.3	36.6	10.6*
n	12.8	1.9*	9.4	48.4*	20.3	13.1	11.9	7.2
l	42.5	44.1	65.0	90.0*	60.3	34.4*	62.8	42.8*

* significant difference between the two positions ($p < .01$).

Tukey range across positions: 7.75% for F12L, 12.63% for F1, 10.38% for F2, 9.88% for F12H.

main effect significant. In contrast the nasals /m, n/ were more intelligible with fewer manner errors in the final position.

In Condition F1, differences in the two positions were found for all dependent variables except voicing errors. Overall intelligibility and the overall perception of place and manner features were most acute for initial consonants. The position-by-consonant interaction was significant for all dependent variables except voicing errors. The consonants /k, f, v, m/ were more intelligible with fewer place errors in the initial position as was the case of the final position for /p, d/. In addition /l/ was more intelligible in the initial position and /z/ showed fewer place errors in the initial position. With respect to manner of production, there were fewer errors for all fricatives, the nasal /n/ and liquid /l/ in the initial position and for /b/ in the final position.

In Condition F2, there were more correct identifications and fewer place and manner errors for consonants in the final position. The position-by-consonant interaction was significant for all dependent variables except voicing errors. The consonants /p, t, d, z/ were more intelligible with fewer place errors in the final position, and intelligibility and the perception of the place feature were more acute for /ð/ in the initial position. In addition, the final position gave more correct identifications for /l/ and fewer place errors for /g, v/. With respect to manner of production, there were fewer manner errors for the voiced plosives, for /f/ and /l/ in the final position and for /ð/ in the initial position.

In Condition F12H, there were more correct identifications and less place errors for consonants in the final position. The position-by-consonant interaction was significant for all variables except voicing errors. There were more correct identifications and fewer place errors in the final position for the plosives /p, t, d, g/ and the fricatives /z, ʃ, ʒ/, and in the initial position for the fricatives /v, ð/. In addition, /m, n, l/ were more intelligible in the final position. The differences noted for manner consisted of fewer errors in the final position for /v, ʒ, m, l/ and in the initial position for /z, ð/. These differences cancelled out the main effect.

In summary the same trends with regard to position were seen for consonant intelligibility and the perception of place in all four conditions. One of the most interesting findings regarding the position effect is that both intelligibility and the perception of place were more acute in the initial position in Condition F1, and in the final position in conditions containing the second formant. Manner errors were fewest in the initial position for Condition F1 and F12L, in the final position for Condition F2, and in Condition F12H the position main effect was cancelled out by some consonants which showed fewer errors in the initial and final positions, respectively.

Comparison of Condition F12L with the unfiltered condition

Table 21 shows the percentages of correct identifications for each consonant for the unfiltered condition as compared to Condition F12L, together with the percentages for each consonant in the two vowel environments and two positions for both conditions. Consonant intelligibility was much higher for the unfiltered material and the largest decreases in intelligibility in Condition F12L, which contained very little energy above the second formant, occurred for /p, t, f, v, s, z, θ, ð, l/. Intelligibility remained relatively high in F12L for /b, d, k, g, ʃ, ʒ, m, n/.

TABLE 21

Consonant intelligibility percentages and percentages for consonants in the two vowel environments and two positions for the unfiltered condition and Condition F12L.

			Vowel				Position			
			Unfiltered		F12L		Unfiltered		F12L	
	Unfiltered	F12L	/I/	/u/	/I/	/u/	Initial	Final	Initial	Final
p	91.6	55.8	95.3	87.5	68.8	42.8	86.6	96.6	25.0	86.6
b	97.7	85.0	99.4	96.0	91.6	78.4	96.3	99.1	86.3	83.8
t	99.7	42.3	100.0	99.4	45.0	39.7	100.0	99.4	31.9	52.8
d	100.0	83.6	100.0	100.0	73.4	93.8	100.0	100.0	71.3	95.9
k	99.4	96.6	99.1	99.7	95.0	98.1	99.1	99.7	95.3	97.8
g	98.0	88.4	98.8	97.2	86.9	90.0	98.5	97.5	83.1	93.8
f	83.1	63.4	91.6	74.7	72.8	54.1	90.3	76.0	64.7	62.2
v	84.1	40.6	85.0	83.1	41.3	40.0	88.5	79.7	54.7	26.6
s	95.6	23.0	95.3	96.0	15.6	30.3	95.7	95.7	23.4	22.5
z	95.3	37.7	92.5	98.2	27.8	47.5	94.7	96.0	20.3	55.0
ʃ	99.2	77.2	99.7	98.8	97.2	57.2	98.8	99.7	76.9	77.5
ʒ	99.2	74.5	99.1	99.4	97.8	51.3	99.7	98.8	71.3	77.8
θ	75.0	31.6	67.8	82.2	20.0	43.1	65.7	84.4	35.0	28.1
ð	69.5	30.2	62.5	76.6	26.3	34.1	73.5	65.6	42.5	17.8
m	98.9	75.2	98.5	99.4	81.6	68.8	99.7	98.2	68.8	81.6
n	99.8	86.6	99.7	100.0	81.6	91.6	99.7	100.0	80.6	92.5
l	100.0	56.7	100.0	100.0	69.1	44.4	100.0	100.0	57.5	55.9

As seen in the vowel-by-consonant interactions of Condition F12L, the consonants /p, f, ʃ, ʒ, l/ were more intelligible in the /I/ environment and /d, s, z, θ, ð/ more intelligible in the /u/ environment. These same trends were only obvious for /p, f, θ, ð/ in the unfiltered material and were particular to the filtered condition for the other consonants.

With respect to consonant position, Condition F12L showed better consonant identification in the final position for /p, t, d, z/ and in the initial position for /v, ð/. These same trends were apparent in the unfiltered material only for /p, v, ð/. In addition, /f/ was better identified in the initial position and /θ/ in the final position for the unfiltered material.

Analysis of Consonant Confusions

The consonant confusions having an incidence of 5% or greater within each condition are listed in Table 22 with the complete confusion matrices for each of the four filter conditions presented in Appendix H.

Examination of Table 22 reveals that the consonant substitutions of Condition F12L involved confusions of only place of articulation for all plosives, all voiceless fricatives and the voiced fricative /z/, and of both place and manner for the remaining voiced fricatives, the nasals and /l/. There were no voicing confusions with an incidence as high as 5%. In Condition F1, there were only place confusions for all voiced plosives, the voiceless plosive /k/ and the nasal /n/, and both place and manner confusions for the remaining voiceless plosives, all fricatives, the nasal /m/ and liquid /l/. As in Condition F12L, there were no voicing confusions reaching 5%. In Condition F2, the substitutions for the voiceless

TABLE 22

Consonant confusions having an incidence of 5% or greater for the four filter conditions.

	Condition F12L	Condition F2	Condition F1	Condition F12H
p	k (28.4), t (10.2)	k (36.9), t (8.3)	t (29.1), k (18.8), f (7.3)	k (33.9)
b	g (6.6)	p (8.8), v (6.7), g (6.1), k (5.3)	d (34.5), g (6.3)	g (5.9), d (5.3)
t	k (26.6), p (25.8)	k (28.1), p (12.8)	p (21.7), k (17.8), f (7.5)	k (36.4), p (18.8)
d	b (13.6)	b (8.9), t (6.6)	b (45.8), g (8.8)	b (17.0)
k	-	-	t (33.1), p (16.9)	t (5.3)
g	d (8.3)	k (22.8), d (11.9)	d (39.4), b (25.5)	d (10.0)
f	θ (20.6)	θ (12.7), s (8.3), v (7.7), m (5.5)	t (18.4), p (13.9), s (8.8), θ (7.8), k (7.2)	θ (13.9), s (6.3)
v	ð (20.8), g (10.0), b (9.5), z (7.3)	ð (18.8), f (6.1), θ (5.9), z (5.6)	z (11.1), n (10.9), b (10.6), d (9.8), g (8.4), ð (6.7)	ð (16.4), g (14.4), b (12.3), l (5.0)
s	f (32.5), θ (32.5), ʃ (5.5)	f (34.4), θ (21.1), ʃ (5.9)	f (29.0), t (10.9), p (9.2), θ (7.0)	f (40.6), θ (28.3), ʃ (10.6)
z	ð (23.0), v (16.3), ʒ (6.4)	ð (20.5), v (19.2), θ (5.3), ʒ (5.2)	v (21.6), g (13.9), d (12.8), l (8.3), ð (6.9)	ð (24.5), v (18.3), ʒ (7.8), n (7.8), l (5.2)

.../contd.

TABLE 22 (contd.)

	Condition F12L	Condition F2	Condition F1	Condition F12H
ʃ	θ (6.9), s (5.8)	ʒ (10.0), s (5.5)	f (32.5), s (15.0), t (10.8), p (8.6), θ (7.5)	θ (8.0), f (7.8)
ʒ	m (6.7), v (6.3), z (5.2)	ʃ (15.5)	v (19.8), d (9.8), m (8.9), g (8.6), b (5.6)	ʃ (12.3), v (5.0)
θ	f (39.2), s (16.3)	f (36.4), s (20.6)	f (27.0), t (21.1), p (14.5), s (11.4)	f (44.7), s (12.3)
ð	v (22.8), z (18.9), n (13.0)	v (29.2), z (23.0), n (8.4), θ (5.2)	v (22.0), z (11.4), b (10.8), d (10.8)	v (21.9), n (13.4), z (13.1), θ (5.5)
m	b (13.2), n (8.3)	n (7.5)	n (35.3), l (11.1), b (6.9), v (6.4)	b (17.2), n (13.3)
n	l (6.3), m (6.1)	m (7.8)	m (30.8)	m (12.8)
l	n (11.6), m (8.0), b (5.6)	n (10.2), ð (9.4), m (8.0)	m (22.0), n (20.6), b (9.8)	n (13.9), m (11.7), v (7.5)

consonants /p, t, s/ and /θ/, and for the nasals /m/ and /n/ involved only place confusions. Those for the voiced consonants /d, g, v, z, ʒ/ and for the voiceless fricative /ʃ/ involved confusions of both place and voicing, and the error responses for /f, b, ð/ and /l/ comprised confusions for all three phonetic features. In Condition F12H there were place confusions only for all plosives, all voiceless fricatives and for the nasal /n/; both place and manner confusions for the voiced fricatives, the nasal /m/ and the liquid /l/; and no voicing confusions reached 5%.

The patterns of confusion among consonants listed in Table 22 can be more clearly visualized by a method of analysis which plots the data in terms of the 'perceptual distance' between consonants along one or more dimensions. Such a spatial representation of the perceptual structure of the 17 consonants was determined for each filter condition by using Kruskal's (1964a) method of non-metric multidimensional scaling, employing the MD-SCAL computer program (Kruskal, 1964b). This method has proved useful in previous analyses of consonant intelligibility under less-than-optimal listening conditions (Shepard, 1972). The multidimensional scaling procedure renders the initially-obscure structure of the confusion matrix more readily apparent, and should lead to a greater understanding of the processing that underlies consonant perception under each filter condition.

Since the MD-SCAL procedure requires that the spatial proximity or perceptual distance between two stimuli be the same in both directions, it was necessary to make the confusion matrices for each condition symmetrical. This was achieved by dividing the total number of confusions for each pair of consonants by the total number of correct responses to these same two stimuli. For example, the entries for the consonants /p/ and /t/ of Table 1 in Appendix H became: $\frac{65 + 165}{357 + 271} = 0.367$ in the symmetrical matrix.

Morgan (1973) recommends that information regarding assymmetry should not be disregarded when interpreting the multidimensional analysis. In the present analysis, therefore, the MD-SCAL results for each condition will be interpreted with reference to the pattern of confusions shown in Table 22.

For a given number of dimensions selected by the experimenter, the MD-SCAL program finds the configuration of the experimental objects (consonants in the present study) that minimizes a measure of fit designated as 'stress' between data and computed values. The stress measure is used not only to find the best configuration for a fixed number of dimensions, but also to determine the number of dimensions most appropriate for the analysis of the data. To the latter end, a plot of stress against the number of dimensions often shows a sharp drop and then a leveling-off or 'elbow' at the number of dimensions that is most appropriate for the analysis (Kruskal, 1964a). Four dimensions were found to be sufficient for initial analysis of the confusion matrices of the present study. The scaling was performed with a maximum of 50 iterations for Conditions F1, F12L and F12H and 100 iterations for Condition F2.

The output of the MD-SCAL computer program consisted of a stress value for each of the best configurations in one, two, three and four dimensional spaces, and a list of the coordinates of each consonant on the dimensions of each best configuration. Figure 5 shows the stress values as a function of the number of dimensions for the four conditions. The most acute 'elbows' in the curves were noted at three dimensions in Condition F12L and at two dimensions in Conditions F1, F2 and F12H. In order to estimate the statistical significance of these results, the stress values were compared to those obtained with random data for a similar number of points and corresponding dimensions (Klahr, 1969). This comparison

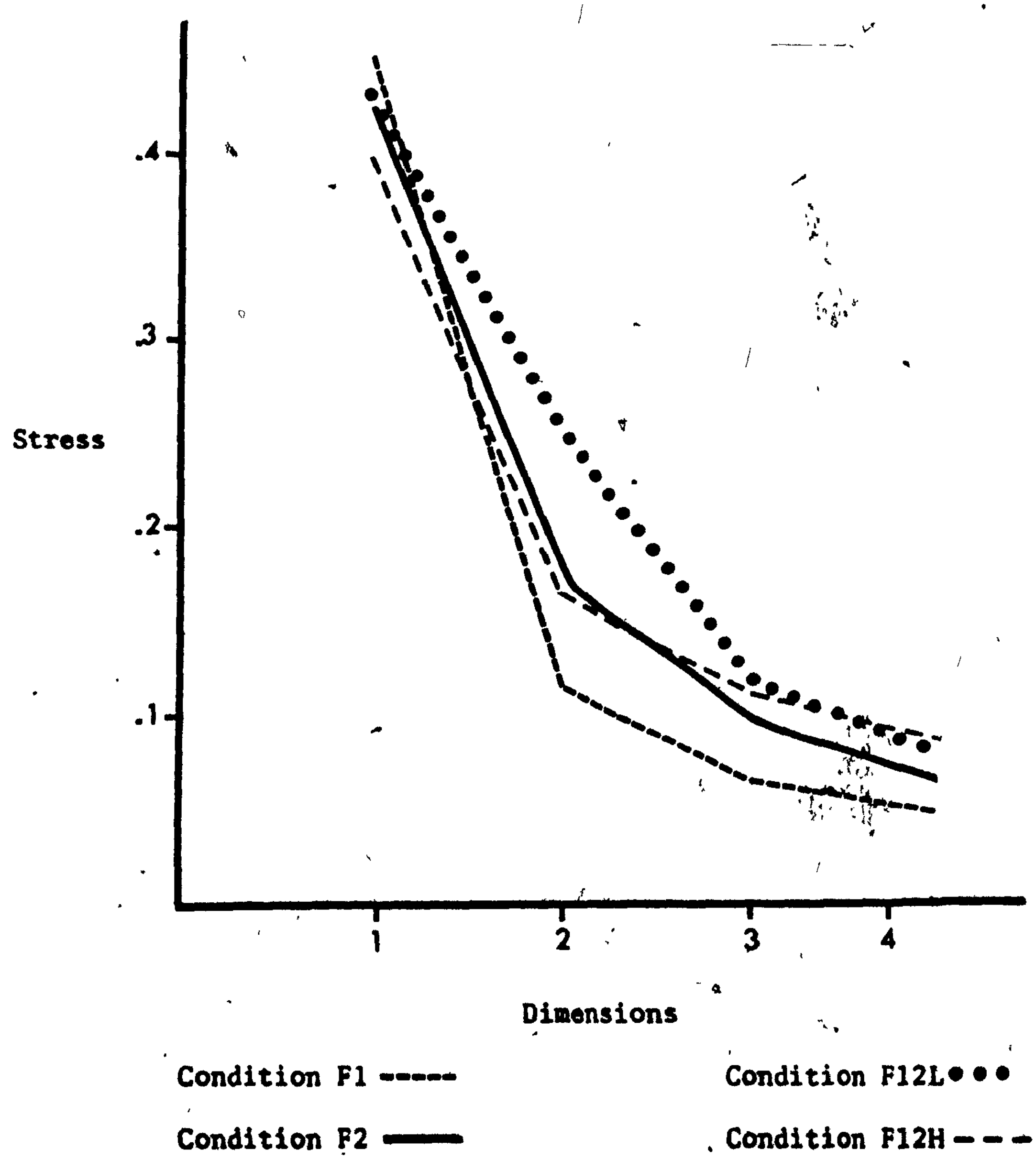

Figure 5: Stress values as a function of one, two, three and four dimensions for the four filter conditions.

indicated that the stress value for the dimensions chosen for each condition would occur less than 5% of the time with random data.

As the initial coordinate system is arbitrary (Kruskal, 1964a), the coordinates obtained for each filter condition were rotated to permit a more meaningful interpretation of the dimensions (Singh and Woods, 1971). For Conditions F1, F2 and F12H, the 17 consonants were plotted with the coordinate values for the two dimensions. The two initially-arbitrary axes were then rotated by hand, keeping the same origin, until the consonant distribution was oriented in a way permitting an interpretation of the dimensions in terms of the three phonetic features. For Condition F12L, in which the elbow was noted at three dimensions, the rotation was performed for two coordinates in the same way as for the other conditions. New coordinate values were then read for each of the 17 consonants on these two dimensions, and these in turn were used with the coordinate values of the third dimension to plot the three-dimensional spatial representation of the consonants. The figures illustrating the perceptual distance for the consonants under each filter condition are based on the rotated coordinates. It should be noted that the rotation of axes and the subsequent naming of dimensions serves merely to enhance the interpretation of the spatially distributed points and to elucidate the relationships of the confusion matrix. The named dimensions have no more 'reality' than the factors derived from factor analysis (Overall, 1964).

Condition F12L

The perceptual distances of the consonants for the three dimensions of Condition F12L are illustrated in Figure 6 where a three-dimensional plot is shown. The three two-dimensional plots from which

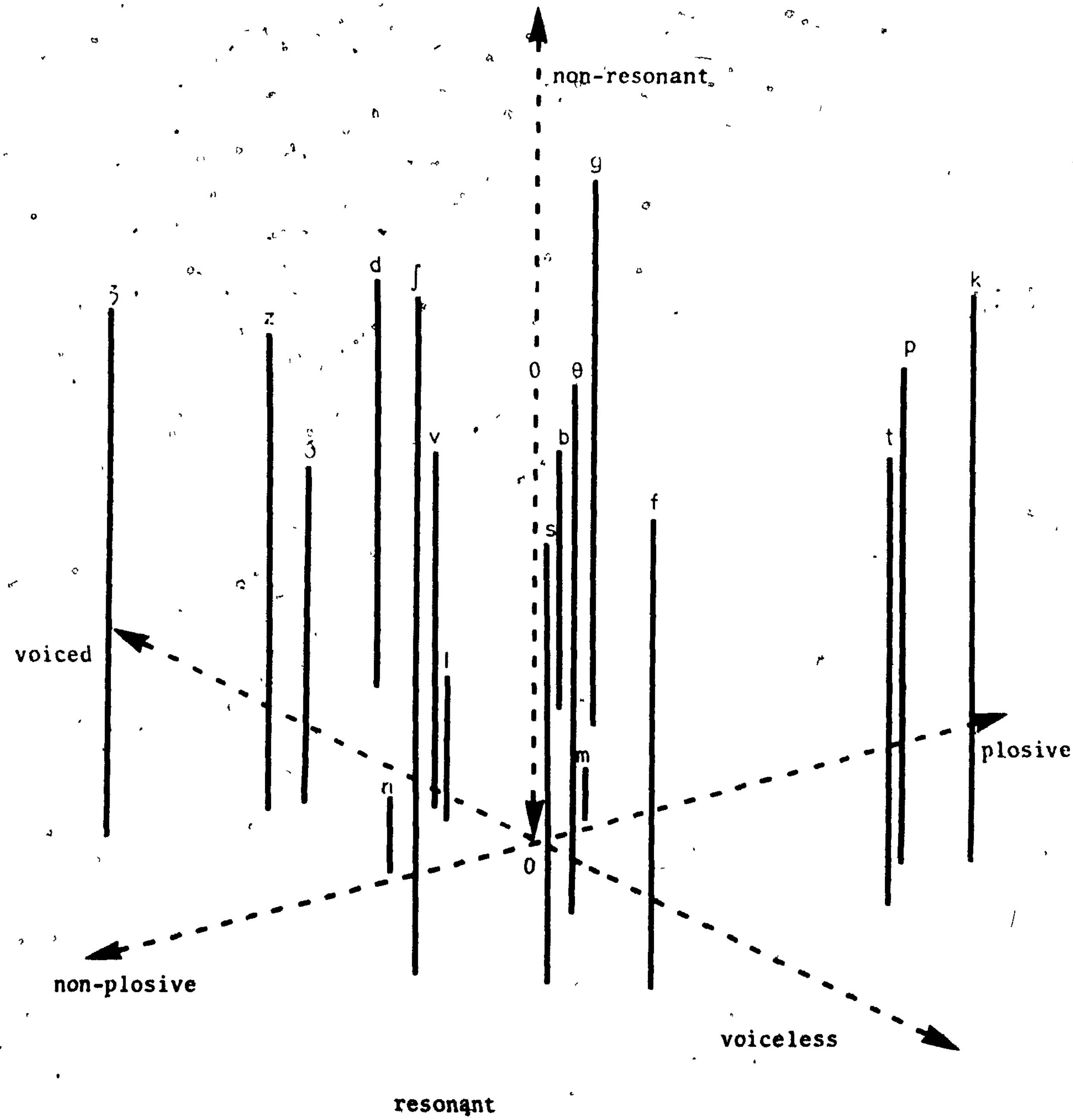

Figure 6: Spatial representation of the 17 consonants for the three dimensions of Condition F12L.

this graph was derived are given in Appendix J. The rotated coordinates partitioned the consonants into voiced and voiceless consonants along one dimension and plosives and non-plosives along a second dimension. Finally, a third dimension designated as resonance and non-resonance separated the consonants /m, n/ and /l/ from the other consonants. Within this three-dimensional plot, small groups of consonants are apparent, namely the nasals /m, n/ and liquid /l/, the voiced plosives /b, d, g/, voiceless plosives /p, t, k/, voiceless fricatives /f, s, ʃ, θ/, and the voiced fricatives /v, z, ʒ, ð/. All groups are formed by consonants sharing the same manner of production and voicing features, with the exception of /l/, which was spatially close to /m/ and /n/. The spatial contiguity of consonants within these groups was, therefore, the result of confusion of place of articulation, indicating, as did the statistical analyses, that place confusions accounted for the majority of errors under this condition. However, the degree of spatial dispersion of the consonants as well as the parallel ordering of the plosives /k, p, t/ and of the fricative /ʃ/ in relation to their voiced cognates /g, b, d/ and /ʒ/, indicates that distinctions of place of articulation were partially preserved in Condition F12L (Shepard, 1972).

The results of the multidimensional analysis shown in Figure 6 clearly illustrate the main confusions under Condition F12L listed in Table 22. No voicing errors occurred with an incidence of 5% or more for any consonant, as is apparent in the well defined voicing dimension found with the MD-SCAL procedure. The main substitutions for all plosives and all voiceless fricatives involved place confusions, as seen by the groups formed by the voiced plosives, voiceless plosives and voiceless fricatives, where the phonemes within each group differed only in place of articulation.

Although manner confusions were relatively infrequent, as indicated by the consonant groupings in Figure 6, some substitutions for the voiced fricative /v/ included manner confusions with the voiced plosives /g/ and /b/, and the spatial proximity of /b/ and /v/ is apparent in Figure 6. Likewise, the most frequent substitution for the nasal /m/ was the voiced plosive /b/, which was closer to /m/ than any other consonant with the exception of /n/ and /l/.

Condition F1

In Condition F1 (Figure 7), the two rotated coordinates partitioned the 17 consonants into those representing the voiced and voiceless consonants along one dimension and those representing the plosives and non-plosives along the other. The voicing distinction was well preserved in this condition, but the plosive distinction was preserved mostly for the voiced sounds. As in the three-dimensional plot for Condition F12L, the groupings of voiced plosives /b, d, g/, voiced fricatives /v, z, ð, ʒ/ and the /m, n, l/ group are clearly evident. However, the voiceless plosive /k, p, t/ and voiceless fricative /f, θ, ʃ, s/ groups, although not overlapping, do not form two separate groups. In fact, /k/ is quite isolated and /p, t/ and /f/ form the only apparent grouping. The formation of the voiced fricative group and the /m, n, l/ group indicates that the features of frication and of resonance were partially retained. In addition, the compactness of the consonant groups illustrates the poor place discrimination in Condition F1.

The results of the multidimensional analysis for Condition F1 can be verified by examining the main consonant substitutions for this condition in Table 22. Only place confusions occurred for the spatially-isolated voiced plosives, demonstrating the correct identification of manner and

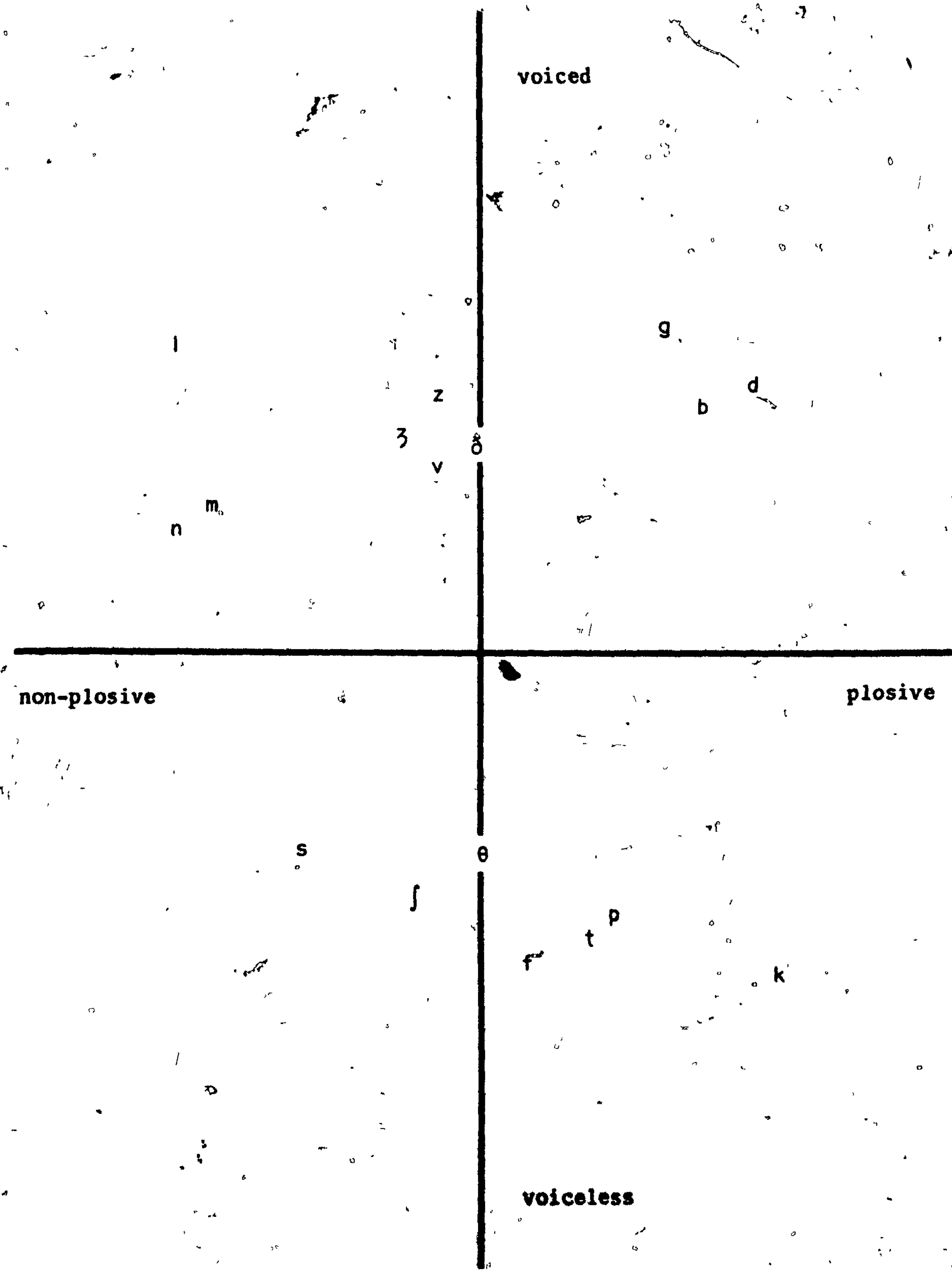

Figure 7: Spatial representation of the 17 consonants for the two dimensions of Condition F1.

voicing in these consonants. The most frequent confusions for the /m, n, l/ group, which was also spatially isolated, were place confusions of /m/ and /n/, as well as /m/ and /n/ confusions with /l/; there were, in addition, relatively infrequent confusions involving the voiced plosive /b/ and the voiced fricative /v/, the reasons for which are not graphically obvious. The voiced fricatives /z, ð, ʒ, v/ which formed a group situated between the voiced plosives and the /m, n, l/ group, showed come confusions involving these two adjacent groups. Except for the spatially-isolated voiceless plosive /k/, the voiceless plosives showed a higher incidence of manner confusions than did their voiced counterparts. The substitutions for the voiceless plosives /p/ and /t/ included a small incidence of /f/ confusions and also represented the most frequent confusions for the latter consonant, as is apparent in the spatial proximity of the three consonants in Figure 7. Although the most frequent errors for the remaining voiceless fricatives /θ, ʃ, s/ were place confusions, some manner confusions with the voiceless plosives also occurred for these consonants. This can be seen graphically by their spatial proximity to the consonants /p/ and /t/.

In summary, there was less separation between voiceless fricatives and plosives in Condition F1 than in F12L, and the dimension separating the resonants /m, n, l/ from the other consonants and more particularly from the voiced fricatives was not as apparent. In addition, the place feature which was only partially present in F12L was even less evident in F1. Otherwise, the features of voicing and 'plosiveness' were quite well preserved in Condition F1.

Condition F2

In Condition F2 (Figure 8), the most appropriate rotation of the axes was not as obvious as in other conditions largely because of the overlapping of the voiced consonants /ʒ/ and /g/ with the unvoiced consonants. The rotation that was finally arrived at divided the consonants into the fricatives and non-fricatives along one dimension and the voiced and voiceless consonants along the other. The voicing dimension was not as clear as in Condition F1, because of the proximity of /ʒ/ and /g/ with the voiceless consonants. Although the grouping of consonants was otherwise less obvious in this condition than in F1, indicating better place discrimination, the following loose groups can be seen: /m, n, l/, /v, ð, z/, /ʒ, ʃ/, /θ, s, f/, /t, k, p, g/ and /b, d/. These groupings indicate the preservation of manner for plosives and of resonance for the /m, n, l/ group. In addition, voicing was invariant within all groups except the groups /ʃ, ʒ/ and /t, k, p, g/ which contained the pairs of consonants /ʃ, ʒ/ and /k, g/ produced in the same manner and articulated at the same place, but opposite in voicing, indicating that the place feature was partially preserved.

The results of the multidimensional analysis for Condition F2 are in agreement with the consonant substitutions (Table 22) observed for this condition. Within the /m, n, l/ grouping, the nasal consonants /m/ and /n/ had only infrequent place confusions with each other; and /l/ was infrequently confused with two adjacent voiced fricatives, as well as with /m/ and /n/. The loosely-grouped voiced fricatives /v, ð, z/ were most often confused with one another, with relatively infrequent confusions involving the nasal /n/ and the voiceless fricatives /θ/ and /f/. Similarly, the voiceless fricative grouping /θ, f, s/ primarily involved internal confusions, with infrequent nasal /m/, voiced fricative /v/ and /ʃ/ confusions.

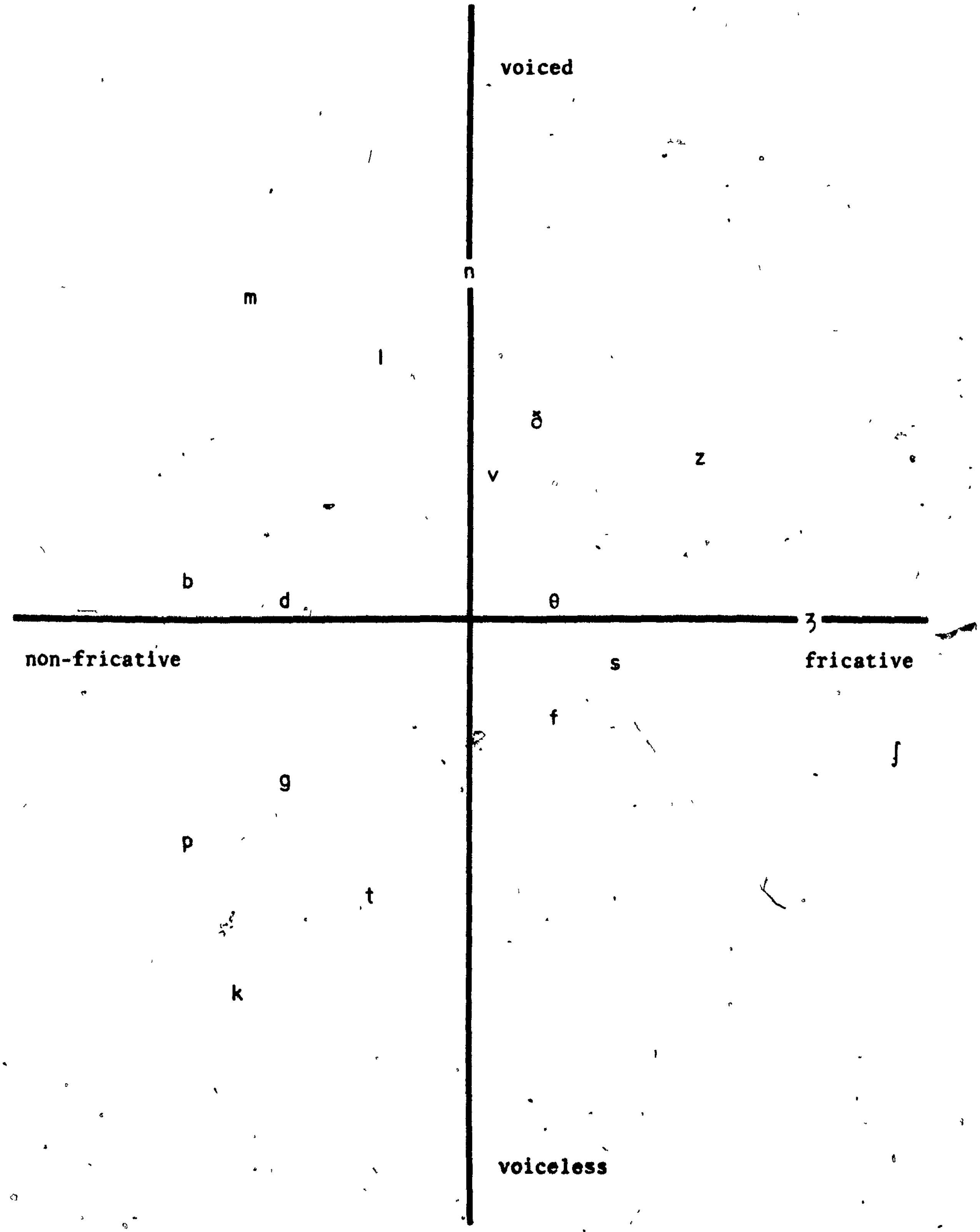

Figure 8: Spatial representation of the 17 consonants for the two dimensions of Condition F2.

The grouped cognates /ʃ/ and /ʒ/ were confused primarily with one another and at low frequency with /s/. Within the plosive group /p, t, k, g/, there were no errors for the voiceless plosive /k/ and the substitutions for the voiceless plosives /p/ and /t/ involved only place confusions; the spatial proximity of the voiced plosive /g/ with these voiceless plosives is reflected by its most frequent substitution being /k/. The relatively infrequent substitutions for the voiced plosive grouping of /b/ and /d/ included both voiced and unvoiced plosives, as well as the voiced fricative /v/, as is apparent from the spatial location of /b/ and /d/.

In summary, the most noticeable difference between Conditions F2 and F12L was a weakening of the voicing distinction. Although plosiveness and resonance appeared as dimensions in Condition F12L, these distinctions were equally preserved in Condition F2. Place was also discriminated about as well in both conditions. A comparison of Conditions F1 and F2 shows a weakening of the voicing distinction and a sharpening of manner distinctions for the fricatives and voiceless plosives, and of the resonance and place distinctions in Condition F2.

Condition F12H

In Condition F12H (Figure 9), the rotated coordinates divided the consonants into voiced and voiceless consonants along one dimension, and plosives and non-plosives along the second dimension. The apparent groupings were the voiced plosives /b, d, g/, the voiceless plosives /p, t, k/ and the voiceless fricatives /s, θ, f, ʃ/. The voiced fricatives /ʒ, z, ð, v/ were not well separated from the resonants /m, n, l/, indicating that the features of frication and resonance were only partially preserved in this condition. The compactness of the plosive and fricative groupings

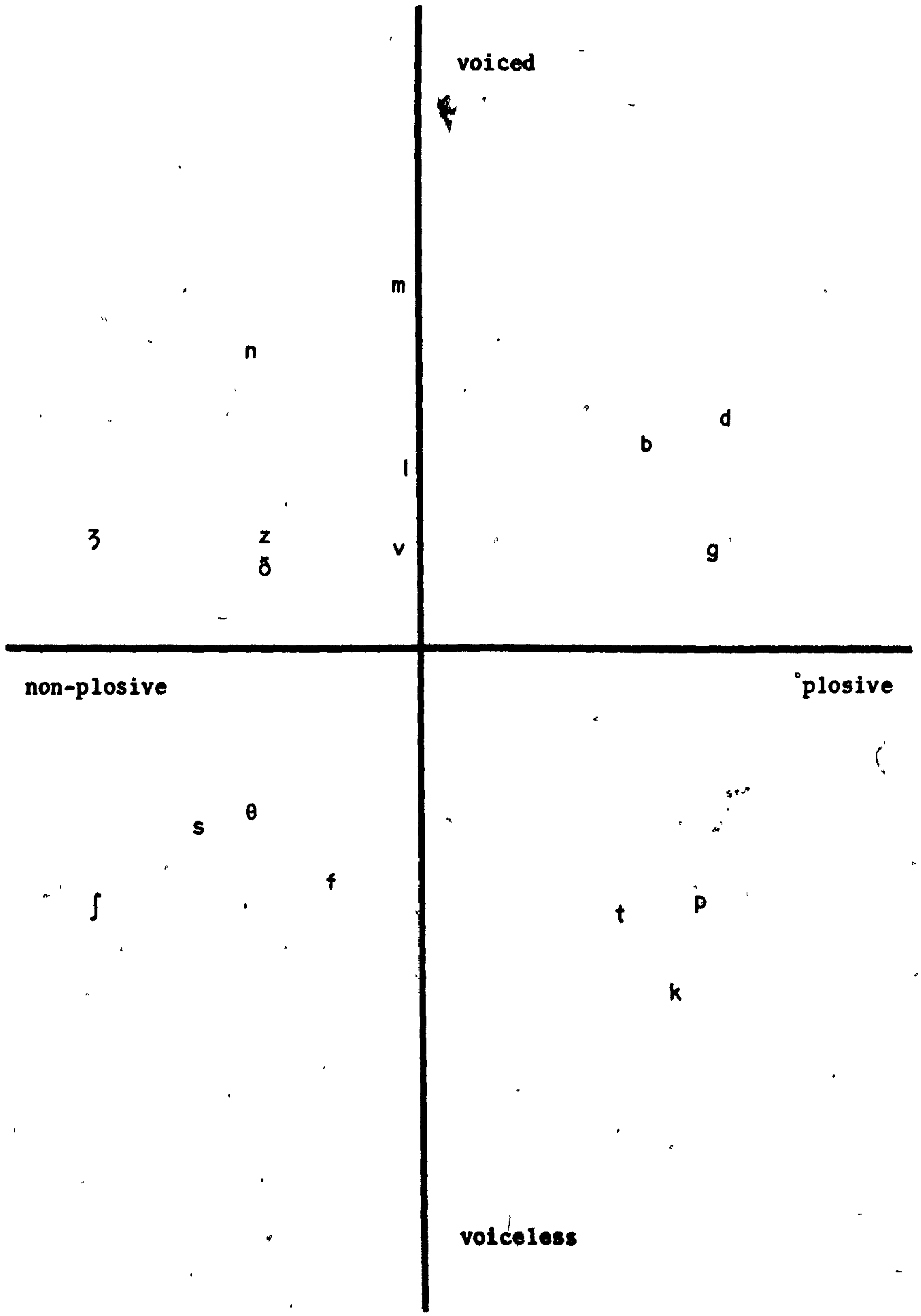

Figure 9: Spatial representation of the 17 consonants for the two dimensions of Condition F12H.

indicates relatively poor discrimination of the place feature.

The confusion matrix results (Table 22) confirm the three definite groups formed by the voiced plosives, voiceless plosives, and voiceless fricatives, since only place confusions occurred within each group. Within the remaining loose grouping, the voiced fricatives were confused with all consonant classes except voiceless plosives, and the resonant consonants were confused with other voiced consonants.

The voicing and plosive distinctions were well preserved in Condition F12H, and the voiceless fricatives were adequately separated but the place distinction was even less apparent and the distinction of voiced fricatives and resonants was greatly weakened as compared with Condition F12L. A comparison of Conditions F12H and F1 shows a sharpening of the voiceless fricative-voiceless plosive distinction along with better place discrimination in F12H. The most noticeable differences between Conditions F2 and F12H were weakening of the fricative distinction and sharpening of the voicing distinctions in F12H. The spatial representation of the consonants was less compact in Condition F12H than in F1, indicating better place discrimination, but more compact than in F2 and F12L, indicating a weakening of place distinctions.

In summary, the main factors underlying consonant distinctions in Condition F12L were voicing, manner of production for plosives and fricatives, the perception of resonance and the partial preservation of the place feature. The absence of the second formant, as seen in Condition F1, resulted in a considerable weakening of the resonant, fricative and place distinctions, but a good preservation of voicing and plosive features. Condition F2 which did not contain the first formant, showed a weakening of the voicing distinction, but the retention of the place feature, manner distinctions for plosives and fricatives and of the resonance feature. The effect of a high level of presentation on both formants, as seen in Condition F12H, resulted in a weakening of the nasal, fricative and place distinctions, with very good preservation of voicing and plosive features.

Consonant substitutions in Condition F12L as compared with the unfiltered condition

Table 23 shows the substitutions occurring with an incidence of 5% or greater for the consonants in Condition F12L and in the unfiltered condition. A much greater number of substitutions occurred for consonants in Condition F12L than in the unfiltered condition. Only the consonants /p, f, v, θ/ and /ð/ showed substitutions occurring with an incidence of 5% or greater for the unfiltered material, while /k/ was the only consonant in Condition F12L for which there were no substitutions equalling 5%. All substitutions for the unfiltered consonants involved place confusions, and the error responses for these consonants also represented the most frequent error responses for these same consonants in Condition F12L.

TABLE 23

Consonant substitutions occurring with an incidence of 5% or greater for Condition F12L and the Unfiltered Condition.

	Unfiltered Condition %	Condition F12L %
p	k (5.2)	k (28.4), t (10.2)
b		g (6.6)
t		k (26.6), p (25.8)
d		b (13.6)
k		-
g		d (8.3)
f	θ (11.7)	θ (20.6)
v	ð (10.2)	ð (20.8), g (10.0), b (9.5), z (7.3)
s		f (32.5), θ (32.5), ʃ (5.5)
z		ð (23.0), v (16.3), ʒ (6.4)
ʃ		θ (6.9), s (5.8)
ʒ		m (6.7), v (6.3), z (5.2)
θ	f (15.3)	f (39.2), s (16.3)
ð	v (17.2), z (5.5)	v (22.8), z (18.9), n (13.0)
m		b (13.2), n (8.3)
n		l (6.3), m (6.1)
l		n (11.6), m (8.0), b (5.6)

DISCUSSION

It was the purpose of this study to assess the contribution of isolated first and second formants to consonant intelligibility and to determine the information content of each formant with respect to the features of place, manner and voicing. The effects of both formants presented at a comfortable level and at a high sound pressure level were also studied. The findings of the present study will be discussed in the following way: 1) the results of the four filter conditions will be compared to the results of other filtered speech studies; 2) the information content of frequencies above the second formant, of the second formant and of the first formant will be related to the important perceptual cues identified in both synthetic and natural speech; 3) the effect of high sound pressure level as seen in Condition F12H will be compared to the effect of high sound pressure level on synthetic formants; 4) the differential effects of vowels and positions will be related to similar findings with natural speech experiments; 5) the perceptual importance of the phonetic features of place, manner and voicing will be evaluated with respect to the findings of the multidimensional scaling procedure; 6) the findings of the present study will be related to current theories of speech perception; and finally, 7) implications of these findings will be discussed with reference to hearing impairment.

Comparison of the Results of the Present Study With Other Filtered Speech Studies

Table 24 summarizes the results of the present study for the four conditions, as compared with the results of other studies using similar filter cut-off frequencies.

It had been hypothesized that intelligibility would be higher for consonants in syllables containing both first and second formants at a comfortable listening level than in syllables containing the second formant alone or in syllables containing both formants at a high sound pressure level. It had also been hypothesized that consonants in syllables containing only the first formant would be the least intelligible. The results of the present study confirmed these hypotheses.

Condition F12L, in which listeners heard both formants at a comfortable level, comprised band-passes ranging in frequency from 130-680 Hz for the first formant and from 800-2200 Hz for the second formant. The intelligibility score of 61.7% agrees with the results (61%) of Franklin (1969a; 1969b), but is slightly lower than the percentages obtained by French and Steinberg (1947), Miller and Nicely (1955), Kryter (1960) and Palva (1965). This discrepancy may be attributed to the shallow rejection slope (24 dB/octave) of the filter used by Miller and Nicely, and to the speech materials used by Palva and Kryter. Both Palva and Kryter used phonetically balanced words which are more intelligible than nonsense syllables (Hirsh, Reynolds and Joseph, 1954). Furthermore, Miller and Nicely, and French and Steinberg used low-pass filters which include all frequencies below a certain frequency, while in the present study the nonsense syllables were presented in two band-pass filters.

TABLE 24

Comparison of results of the present study with other filtered speech studies.

Study	Filter Cut-Off Frequencies (Hz)	Filter Rejection Slope (dB/octave)	Speech Material	Intelligibility
CONDITION F12L				
Present study	130-680 800-2200	72	CV & VC syllables	61.7%
French & Steinberg (1947)	low-pass: 1950	Not given	Nonsense syllables	68.0%
Miller & Nicely (1955)	low-pass: 2000	24	CV syllables	68.0%
Kryter (1960	250-750 1250-1750	70	PB words	72.0%
Palva (1965)	480-720 1800-2400	60	Finnish PB words	78.3%
Franklin (1969a, 1969b)	240-480 1020-2040	60-70	Fairbank's Rhyme Test	61.0%
CONDITION F1				
Present study	130-680	72	CV & VC syllables	23.2%
French & Steinberg (1947)	low-pass: 750	Not given	Nonsense syllables	15.0%

TABLE 24 (contd.)

Study	Filter Cut-Off Frequencies (Hz)	Filter Rejection Slope (dB/octave)	Speech Material	Intelligibility
CONDITION F1 (contd.)				
Miller & Nicely (1955)	low-pass: 600	24	CV syllables	48.0%
Kryter (1960)	0-500	70	PB words	13.0%
	500-1000	70	PB words	32.0%
Palva (1965)	480-720	60	Finnish PB words	25.7 ± 10.4%
Franklin (1969a, 1969b)	240-480	60-70	Fairbank's Rhyme Test	Unintelligible
CONDITION F2				
Present study	800-2200	72	CV & VC syllables	56.0%
Kryter (1960)	1250-1750	70	PB words	37.0%
	1500-2000	70	PB words	36.0%
Palva (1965)	900-1800	60	Finnish PB words	52.0 ± 12.6%
Franklin (1969a, 1969b)	1020-2040	60-70	Fairbank's Rhyme Test	40.0%

TABLE 24 (contd.)

Study	Filter Cut-Off Frequencies (Hz)	Filter Rejection Slope (dB/octave)	Speech Material	Intelligibility
CONDITION F12H				
Present study	130-680 (100 dB SPL) 800-2200 (80 & 84 dB SPL)	72	CV & VC syllables	56.4%
French & Steinberg (1947)	low-pass: 1950 (30 dB orthotelephonic gain (1))	Not given	Nonsense syllables	65.0%
Franklin (1969a, 1969b)	240-480 (40 dB SL) 1020-2040 (0 dB SL)	60-70	Fairbank's Rhyme Test	38.0%

(1) "By definition a telephone system has an orthotelephonic response of 0 dB when it can be replaced by a one meter air path, between talker and listener without changing the loudness of the received speech at any frequency." French & Steinberg (1947), pp. 98.

The results obtained for Condition F1 (23.2%) are lower than those obtained by Miller and Nicely (48%). This discrepancy can again be attributed to the shallow rejection slope used by Miller and Nicely. With respect to other studies, the results of the present investigation are higher than those obtained by Kryter with his 0-500 Hz band, by French and Steinberg, and by Franklin. It is difficult to accurately make comparisons with the results of Kryter because of the fixed 500 Hz bandwidths used in his study, and with French and Steinberg because the rejection slope of the filters is not reported. Franklin only reported that her low 240-480 Hz band was 'unintelligible'. The results of the present study for Condition F1 agree with those of Palva.

The results for Condition F2 (56.0%) are also in good agreement with those of Palva (1965) who used a filter with a similar rejection slope, but are higher than those obtained by Kryter (1960) with a fixed 500 Hz bandwidth. With respect to Franklin's study, the disagreement may lie in the levels of presentation. In the present study, Condition F2 was presented at 46 and 50 dB SPL while Franklin presented her high band at threshold (0 dB SL).

Condition F12H showed an overall decrease of 5.3% in intelligibility as compared with Condition F12L. French and Steinberg (1947) found a decrease of 3% at 30 dB orthotelephonic gain relative to 0 dB orthotelephonic gain. The decrease noted by Franklin was more considerable owing to the respective levels of presentation of the two bands. Her high band was presented at threshold and the low band at 40 dB SL. These levels do not correspond directly to a high level of presentation, but show the decrease in intelligibility which occurs when the low band is more intense than the high. In the present study, the same relative difference between formant intensities was maintained in Conditions F12H and F12L.

This comparison of the results of the present study with the results of other filtered speech studies demonstrates that similar percentages of overall intelligibility were obtained relative to similar filter cut-off frequencies, regardless of the speech materials used. The only study which would have permitted a direct comparison with the present study with respect to formant frequencies is that of Miller and Nicely who used consonants with one vowel (/ɑ/). Because of the shallow rejection slopes of their filters, this comparison becomes impossible.

The Information Content of Frequencies Above the Second Formant, and of the Frequencies of the Second and First Formants

As the band-pass cut-off frequencies of the filter used in the present study were fixed for the entire duration of the syllables, it is necessary, when discussing the results of individual consonants, to take into account the invariant energy included by each band-pass filter. A more elegant method would have extracted the second formant and second formant transition of the vowel together with the corresponding region in the invariant energy (Atal and Hanauer, 1971; Olive, 1971). The present method resulted in a wider bandwidth of the filtered signal than would have been obtained with a tracking technique, and consequently more invariant energy was included in the filtered utterance. If all invariant energy had been eliminated in the syllables prior to filtering, this would have modified the durational characteristics of the sounds and reduced the four classes of consonants to a class similar to plosives (Gerstman, 1956; Barrs, 1963; Cole and Scott, 1973b).

Frequencies above the second formant

Place of articulation. It had been hypothesized that Condition F12L would be the most intelligible filter condition and that the smallest number of place errors would occur in Conditions F12L and F2. These hypotheses were confirmed. A comparison of place confusions in Condition F12L with those of the unfiltered condition (Table 23, p. 111) indicated that the second formant frequencies did not contain all the necessary place information. This is also apparent in the number of place errors in Conditions F12L (33.9%) and F2 (35.0%) (Table 8, p. 68). These results agree with Miller and Nicely (1955) who found that the elimination of frequencies above 1900 Hz produced a deterioration in the perception of the place feature.

When the frequencies above the second formant were eliminated, the least intelligible consonants were /p, t, f, v, s, z, θ, ð, l/ (Table 21, p. 92), which also showed the highest number of place errors (Table 23, p.111). The increase in place errors in Condition F12L relative to the unfiltered condition may be attributed to the absence of the third formant transition (Harris et al., 1958; Hoffman, 1958; Delattre, Liberman and Cooper, 1964) and to the absence of invariant energy above the second formant. The two vowels used in the present study were a front vowel /I/ which has a relatively intense third formant, and a back vowel /u/ which has a weak third formant (Peterson and Barney, 1952). The importance of the third formant transition will be assessed when discussing the vowel effects.

The large increases in place errors for /p/ and /t/ when the frequencies above the second formant were absent indicate that the second formant transition was not a sufficient cue for the accurate identification of these plosives. Cooper et al. (1952) and Liberman et al. (1954) stated that second formant transitions served as cues for aurally perceived distinctions among both voiced and voiceless stops. In pattern playback studies, the only distinction between the two categories of consonants is the extent of first formant transition, the second formant transitions being equally represented for voiced and voiceless stops. In natural speech, however, very little second formant transition is seen after /p, t, k/, as most movements of the speech organs from the consonant to the vowel take place during the aspiration phase (Fischer-Jorgensen, 1972a). The present results suggest that the invariant energy above 2200 Hz, which was the highest cut-off frequency of the second formant band-passes, is important for the identification of /p/ and /t/. The importance of bursts for voiceless plosives is well documented in the literature on natural speech (Halle, Hughes and Radley, 1957; Malécot, 1958; Wang, 1959; Ainsworth, 1968a; Cole and Scott, 1972; 1973a; Fischer-Jorgensen, 1972d; Scheib and Winitz, 1972; Winitz, Scheib and Reeds, 1972). The present results are in agreement with Liberman et al. (1952) and Halle et al. (1957) for /t/, but in disagreement for /p/, which has been described as having a burst concentrated in the low frequencies. This suggests that either the high frequency aspiration noise is important for the identification of /p/ (Fischer-Jorgensen, 1972c) or the burst for /p/ is not concentrated, but neutral, having an even distribution of energy at all frequencies (Fischer-Jorgensen, 1954). When both formants were present at a comfortable level, the most frequent substitutions for /p/ and /t/ were /k/. Similar results were obtained by Fant (1968) who low-pass filtered speech at 2000 Hz and found that bursts from both /t/ and /p/ became more concentrated and similar to /k/. The second formant transition is thus not a sufficient representation of relevant acoustic information for these two phonemes and Liberman et al's (1967) conclusions in this regard are not necessarily valid for real speech.

The fricatives /f, v, s, z, θ, ð/ were also among the least intelligible consonants (Table 21, p. 92) with the highest number of place errors (Table 23, p. 111) when the frequencies above the second formant were absent. Transitions have been shown to be important in differentiating /f/ from /θ/, and /v/ from /ð/, and noise important for making /s-ʃ/ and /z-ʒ/ distinctions (Harris, 1954; 1958; Heinz and Stevens, 1961; Cole and Scott, 1973b). Both /s/ and /z/ have their main concentrations of energy above 3600 Hz (Harris, 1956; Hughes and Halle, 1956; Strevens, 1960; Jassem, 1965), and the increase in place errors when only the first two formants were present can be explained by the elimination of these frequencies. The most frequent substitutions for /s/ and /z/ under these circumstances were the fricatives /f/ and /ð/, respectively, which have concentrations of energy in the mid frequencies (Jassem, 1965).

Although the main concentrations of energy for /f, v, θ, ð/ were included in the second formant band-passes, these sounds were confused with one another within the same voicing category. The second formant transition and the invariant energy included in the band-passes were therefore not sufficient acoustic cues. These four fricatives were also the least intelligible consonants in the unfiltered condition, indicating, as found by Sharf (1968), that they are generally weak in acoustic cues.

Manner of production. The elimination of the frequencies above the second formant did not markedly affect the perception of manner (Table 8, p.68), which is in agreement with the results of Miller and Nicely (1955) and of Gay (1970). An analysis of the confusions for the individual consonants revealed that the voiced fricatives showed a slight increase in manner confusions in the absence of high frequencies (Table 23, p. 111). Manner cues for fricatives have been attributed to invariant energy (Liberman et al., 1959; Fant, 1960; Delattre et al., 1964) and to the first formant transition (Delattre, 1958). The present results suggest that the invariant energy above 2200 Hz contains some manner information for the voiced fricatives. Blesser (1972) found, as in the present study, that voiced fricatives were often perceived as nasals in speech which had been filtered with a 200-3000 Hz band-pass, prior to spectral transformation. The absence of frequencies above 3000 Hz or 2200 Hz causes the low frequency component of voiced fricatives (Hughes and Halle, 1956) to be perceived as a nasal resonance. The perception of voiced fricatives as voiced plosives in the absence of high frequencies may also be due to the perception of the low frequency band and to the shorter duration of voiced fricatives than voiceless (Denes, 1955).

The liquid /l/ showed an increase in manner errors in the absence of high frequencies. This consonant was at a disadvantage, however, since it was the only consonant of its class, and any error response for /l/ constituted a manner error. The important manner cues for liquids are the steady-state formants, which remain at the loci for 30-50 msec. (O'Connor et al., 1957), the first formant transition (Delattre, 1958) and the slow transitions which distinguish nasals from liquids (Cooper et al., 1952; O'Connor et al., 1957). The most frequent substitutions for /l/ were the nasals indicating that the higher resonances may in fact contribute to the nasal-liquid distinction.

Voicing. The perception of the voicing feature was not markedly affected by the absence of frequencies above the second formant. These results are again in agreement with those of Miller and Nicely (1955) and Gay (1970).

The second formant frequencies

It had been hypothesized that the filter condition which did not contain the second formant would show the lowest intelligibility and the highest number of place errors. It had further been hypothesized that removal of the second formant would result in an increase in manner and voicing errors, since it was thought that both features would be best perceived when both formants were present. The results of the present study confirmed these hypotheses. When the second formant was eliminated, intelligibility decreased from 61.7% in Condition F12L to 23.2% in Condition F1, and the perception of place, manner and voicing became less acute (Table 8, p. 68). These findings were apparent in the more compact multidimensional spatial distribution of the consonants, which also revealed a reduction in consonant class distinctions in Condition F1 relative to Condition F12L (Figure 7, p. 103).

The information content of the second formant for each consonant can be derived by comparing the results of Conditions F12L and F1 to determine the information lost by the elimination of the second formant frequencies; by comparing Conditions F1 and F2 to determine the relative information contributed by the first and second formant; and by studying the differences between consonants within Condition F2 to assess the extent to which consonants can be identified when only second formant frequencies are heard.

Place of articulation. The results of removing the second formant frequencies, as seen by comparing Conditions F12L and F1, indicated that the second formant frequencies contained place information for all consonants except /t/ and /z/. Similarly, the comparison of Conditions F1 and F2 indicated that the second formant contained more place information than the first for all consonants except /t, s, z/. As was discussed in the preceding section, these three consonants depend on high frequency energy above the second formant for identification.

The place information contained by the second formant frequencies must be discussed with respect to the invariant energy included in the band-pass filter and to the importance of the second formant transition. The results of the Tukey test of honestly significant differences between consonants in Condition F2 showed that the second formant frequencies contained more information for some consonants than for others (Table 10, p. 74). The most intelligible consonants with the smallest number of place errors were those characterized by second formant transitions and by invariant energy included by the band-pass filter. This was the case for /k/ which has a burst in the mid frequencies (Liberman et al., 1952; Fischer-Jorgensen, 1954; Halle et al., 1957), for /ʃ/ and /ʒ/ which have concentrations of friction noises in the mid frequencies (Jassem, 1965), for the nasals and for the voiced plosives. Although some place information was contained in the second formant frequencies for /f, v, t, p, s, z, ð, θ/, the low intelligibility and high number of place errors for these consonants indicated that they depended on high frequency invariant energy for identification (Table 10, p. 74).

A comparison of the results for voiced and voiceless plosives indicates that the place information content of second formant frequencies was not the same for both categories of stops. Voiceless plosives depend strongly on noise bursts (Ainsworth, 1968a; Fischer-Jorgensen, 1972d) and the only burst included in the band-pass filter was the burst of /k/. The second formant transition was not as sufficient a cue for /g/ as was the burst for /k/, while the transitions for /d/ and /b/ enabled a more accurate perception of place than the transitions for /t/ and /p/. These results show that the importance of the second formant transition is not the same for voiced and voiceless plosives in contrast to the results of pattern-playback studies (Cooper et al., 1952; Liberman et al., 1954).

Manner of production. With the second formant removed as seen in Condition F1, manner errors increased for all consonants except /d/ and /g/, indicating that the second formant frequencies contained manner information for most consonants. This was also apparent in the comparison of Conditions F2 and F1 which showed that Condition F2 contained more manner information than F1 for all fricatives, nasals, /l/ and the voiceless plosives /p, t/. Both conditions contained equal information for /k/ and /g/. Manner cues at the level of the second formant frequencies are found in the invariant energy contained by the band-pass filters and possibly in the duration of transitions (Liberman, Delattre, Gerstman and Cooper, 1956). The consonants which were most affected by the absence of second formant frequencies were the fricatives. Transition duration has not been studied in the differentiation of fricatives from other classes of consonants, but since the elimination of invariant energy reduces the class of fricatives to plosives (Gerstman, 1956; Barrs, 1963; Cole and Scott, 1973b), this suggests that transition duration is a very weak cue for manner in these consonants. The present results show that manner information was contained for fricatives in the invariant energy at the level of the second formant frequencies. The results also suggest that there is very little manner information for fricatives in the first formant transition.

Although the increase in manner errors was most noticeable for the fricatives, the results for the voiceless plosives /p/ and /t/ also indicated that the invariant energy at the level of the second formant contributes to manner distinctions. Both the first and second formants contained the same amount of manner information for /k/ and /g/, this information being contained in the burst and temporal aspects of the invariant energy in the second formant frequencies.

The increase in manner errors for the nasals in the absence of the second formant frequencies may be attributed to the elimination of the second nasal resonances and also perhaps to the absence of transition rate of the second formant (Cooper et al., 1952; O'Connor et al., 1957). As more manner information was contained in the second formant frequencies than in the first, the results suggest that these cues are perceptually more important than the first nasal formant (Fujimura, 1962), and than the absence of a first formant transition (Liberman et al., 1954). Similarly, the steady-state formant remaining at the locus and the continuity of the second formant appear to be the main contributors of manner information for /l/ (Cooper et al., 1952; O'Connor et al., 1957).

When the second formant frequencies were absent, the substitutions for each fricative included nasal or plosive responses, clearly indicating the importance of these frequencies for making fricative distinctions. On the other hand, the plosives /p/ and /t/ showed a small incidence of /f/ responses, and the nasal /n/ was consistently confused with /m/, indicating that manner information for these consonants was not exclusively contained in the second formant frequencies.

Voicing. The effect of the absence of the second formant on the perception of the voicing feature is shown by the increase in voicing errors for the voiceless fricatives /f, s, ʃ/ and the voiced consonants /b, ʒ, m/ in Condition F1. The important cues for voicing at the level of the second formant frequencies are the voice onset time for plosives (Abramson and Lisker, 1965; Lisker and Abramson, 1967b), the noise during the second formant transition for voiceless plosives (Liberman, Delattre and Cooper, 1958), and the frequency and durational aspects of the invariant energy for fricatives (Denes, 1955). The increase in voicing errors for /m/ is related to an increase in manner errors. Although phonetic features are perceived almost independently of one another (Miller and Nicely, 1955), it has been shown that under certain conditions of distortion, interaction may occur between two features (Pickett and Rubenstein, 1960). This may explain the increase in voicing errors for /m/.

A comparison of Conditions F1 and F2 for voicing errors indicated that the second formant frequencies contained less voicing information than the first formant for the voiced plosives and for the voiced fricatives, but more information for /s/ and /m/. The voicing information for /s/ was contained in invariant energy and the perception of voicing for /m/ was related to the perception of manner. The increase in voicing errors in Condition F1 was not apparent in the spatial representation of the consonants (Figure 7, p. 103), nor in the pattern of consonant confusions (Table 22, p. 94). These results indicate that although the overall number of voicing errors increased when the second formant frequencies were eliminated, some of the information contained by the second formant was also present in the first formant frequencies.

The first formant frequencies

It had been hypothesized that absence of the first formant frequencies would not increase the number of place errors, but would decrease overall intelligibility and the perception of manner and voicing. The results of the present study confirmed the hypothesis for intelligibility, place and voicing, but not the hypothesis for manner. Intelligibility changed from 61.7% in Condition F12L to 56.0% in Condition F2. The number of place and manner errors were similar in both conditions, although it had been hypothesized that manner would be best perceived when both formants were present. Voicing errors increased from 3.9% in Condition F12L to 11.2% in Condition F2 (Table 8, p. 68). These results were apparent in the spatial representation of the consonants in Condition F2, which showed good preservation of place and manner distinctions, but a weakening of the voicing dimension (Figure 8, p. 106).

The same approach was used to assess the information content of the first formant frequencies as was used to determine the content of the second formant frequencies. Comparisons were made between Conditions F12L and F2, and between F1 and F2, and the differences between consonants in Condition F1 were also examined.

Place of articulation. A comparison of individual consonants indicated that although the absence of the first formant frequencies did not alter the perception of place for the majority of consonants, place errors increased slightly for /b/ and /g/. This might be explained by the fact that the burst for /b/ contains energy throughout the frequency range (Fischer-Jorgensen, 1954) and the removal of the first formant frequencies was eliminating part of the burst. A more likely explanation is that the weakening of the perception of the other features influenced the perception of place in these consonants (Pickett and Rubenstein, 1960). A comparison of Conditions F12L and F2 showed that the elimination of the first formant improved the perception of place in the voiced fricatives /v/ and /ʒ/ which was probably related to the improved manner discrimination for these sounds in the absence of the low frequencies.

Manner of production. Although it had been hypothesized that the perception of manner would deteriorate in the absence of the first formant, this hypothesis was not confirmed. Manner cues contained in the first formant frequencies are found in the first formant transition (Delattre, 1958), and in the low frequency invariant energy such as nasal resonances (Fujimura, 1962). When the first formant frequencies were absent, manner errors increased only for /b/, /d/ and /n/. The filtering out of low frequencies eliminated the low frequency nasal resonances which did not affect the perception of manner for /m/, and although manner errors increased for /n/, this was not apparent in the confusions for this consonant (Table 22, p. 94). A comparison of Conditions F1 and F2 indicated that more manner information was contained in the first than in the second formant only for /b/ and /d/ and that the first and second formants contained equal manner information for /k/ and /g/. These results indicate that the first formant transition is an important cue for plosives and in particular voiced plosives, and are in agreement with the results of Delattre et al. (1955), Stevens and House (1956) and Delattre (1958; 1962).

The spatial representation of the consonants in Conditions F12L and F2 indicated that although manner errors increased for /b/ and /d/ when the first formant frequencies were absent, the distinction separating plosives from the other classes of consonants was very well preserved (Figure 6, p. 100; Figure 8, p. 106). This is also seen in the only manner confusion for plosives which was the small percentage of /v/ responses for /b/.

The elimination of the first formant improved manner perception for /v/ and although the result was not significant, a similar trend was seen for /ʒ/. When only the second formant frequencies were present, these two phonemes were confused with fricatives, while they showed plosive and nasal confusions in Condition F12L. This indicates, as found by Blesser (1972), that the presence of the low frequency voicing resonance interferes with the perception of manner when higher frequencies are eliminated.

Voicing. The marked weakening of the voicing distinction in the absence of the first formant frequencies was apparent in the increase in voicing errors for the voiced plosives and for /f/, /v/ and /ʒ/ in Condition F2. In addition, a comparison of Conditions F1 and F2 indicated that the first formant frequencies contained more voicing information than the second formant frequencies for the voiced plosives and for the voiced fricatives. The important cues for voicing contained in the first formant frequencies are the voice bar and more particularly the first formant transition for plosives (Cooper et al., 1952; Liberman, Delattre, Cooper and Gerstman, 1954; Delattre et al., 1955; Liberman, Delattre and Cooper, 1958; Stevens and Klatt, 1971; Haggard and Summerfield, 1972; Summerfield and Haggard, 1972), and the voice bar and low intensity formants running through the friction for voiced fricatives (Hughes and Halle, 1956; Delattre, Liberman and Cooper, 1964).

The number of voicing errors in both Conditions F1 and F2 was relatively low, probably because voicing identification involves only a binary decision and also because acoustically, voicing is well represented at the level of both first and second formants. An interesting finding is that although both conditions did not differ appreciably with respect to the number of overall voicing errors, they did in fact differ with respect to the perceptual saliency of the feature as seen by the weakening of the voicing dimension in Condition F2. This was apparent in the spatial representation of the consonants and in the increase in voicing confusions for that condition. The most frequent responses in Condition F2 for /b/, /g/ and /ʒ/ were their voiceless cognates. The /ʒ/ response for /ʃ/, and /v/ and /m/ responses for /f/ cannot be adequately explained, but may be the result of distortion occurring in the filtering process.

Effect of High Sound Pressure Level

In view of the work of Martin, Pickett and Colten (1972), Martin, Osberger and Pickett (1972), Danaher, Osberger and Pickett (1973) and Pickett and Danaher (1973), it had been hypothesized that a decrease in intelligibility and an increase in place errors would occur in Condition F12H relative to Condition F12L. The results of the present study confirmed the hypothesis as illustrated by the more compact spatial representation of the consonants in Condition F12H (Figure 9, p. 108).

Since Martin, Osberger and Pickett (1972) had found that at a high sound pressure level the presence of the first formant interfered with the perception of second formant transitions, it was thought that those consonants cued by second formant transitions would show the largest increase in place errors in Condition F12H relative to Condition F12L. An analysis of consonants (Table 9, p. 70 and Appendix F) revealed that no individual consonant showed a significant increase in place errors, although the overall number of place errors in Condition F12H was significantly greater than in Condition F12L. These results may be explained by the fact that the second formant transitions were not the only place cues present, but that the invariant energy included by the band-pass filters also contributed to consonant identification. It had also been hypothesized that a greater increase in place errors would occur in consonants in the /u/ environment than in the /I/ environment in Condition F12H because of the frequency proximity of the first two formants in this vowel. This hypothesis was not substantiated (Table 16, p. 84). Although frequency proximity had been stated as an important factor in the spread of masking of second formant transitions by first formants, Danaher, Osberger and Pickett (1973) later showed that the relative amplitude of the first and second formants had a greater effect on discrimination than the proximity of the formants. The two vowels /I/ and /u/ used in the present study, had similar relative amplitudes of their first two formants (Peterson and Barney, 1952), and therefore the results for the two vowels are not surprising.

An unexpected finding was the increase in manner errors for /v, z, l/ in Condition F12H, which is illustrated by the spatial proximity of the resonants and voiced fricatives in Figure 9 (p.108). This trend was also apparent for the voiced fricatives in Condition F12L, and is due to the presence of the low frequency resonances in the invariant energy and to the absence of the higher frequencies (Blesser, 1972).

Vowel and Position Effects

Vowels

The two vowels /I/ and /u/ used in the present study were primarily selected because they enabled an adequate attenuation of unwanted formants by band-pass filtering. As the results of natural speech studies had indicated that the perceptual importance of formant transitions was related to the magnitude of formant movements (Andresen, 1960; Ainsworth, 1968a; Fischer-Jorgensen, 1972a; 1972b; 1972c), it was hypothesized that for all conditions containing the second formant, place errors and therefore intelligibility would vary as a function of vowels. This hypothesis was confirmed.

Table 25 shows the vowels which gave the smallest number of place errors and highest intelligibility scores for the consonants in the three conditions containing the second formant. The differences in intelligibility and in the perception of the place feature for the consonants in the two vowel environments can be explained as follows:

- 1. The vowels /I/ and /u/ represent a front and a back vowel and the extent of the transition depends on the place of articulation of the adjoining consonant relative to the vowel. In addition, /I/ is a lax vowel while /u/ is tense. Lax vowels have been described as having short targets and long transitions while tense vowels have longer targets and relatively shorter transitions (Lehiste and Peterson, 1961).

- 2. The vowels /I/ and /u/ differ in the relative amplitudes of their third formants, /I/ having a relatively intense third formant while the third formant for /u/ is relatively weak (Peterson and Barney, 1952).

TABLE 25

Vowels giving the smallest number of place errors and highest intelligibility scores for the consonants in the three conditions containing the second formant.

	Condition		
	F12L	F2	F12H
p	I*	I	I
b			
t			
d	u	u	u
k			
g			
f	I*		I
v			
s	u	u	
z	u*		
ʃ	I	I	I
ʒ	I	I	I
θ	u*		u
ð	*		u
m			
n			
l			

* Trend also apparent in the unfiltered material.

N.B. Blanks indicate that the differences in scores for the consonants in the two vowel environments were not significant.

Third formant transitions have been shown to be important cues for place perception for some consonants in front vowel environments (Harris et al, 1958; Delattre, Liberman and Cooper, 1964). In the present study, the absence of perceptually important third formant transitions would be expected to show a greater effect for consonants in the /I/ environment than in the /u/ environment.

- 3. The vowels /I/ and /u/ differ in the frequency of their second formants, /I/ having a higher second formant than /u/ (Peterson and Barney, 1952). As a result of the method used in the present study, the filtering of consonants in the /I/ environment may have included more invariant energy than for consonants in the /u/ environment. Table 26 shows the high cut-off frequencies of the second formant band-passes for the groups of consonants in the two vowel environments.

The difference in the extent of the transitions for the two vowel environments can explain the more acute perception of place and therefore intelligibility for /p/ with /I/ and /d/ with /u/. These results are in agreement with Halle, Hughes and Radley (1957), Harris et al (1958), Wang and Fillmore (1961), Ainsworth (1968a), and Fischer-Jorgensen (1972a), and have been explained by the small transitions of bilabial plosives with /u/ and of alveolars with /I/. The perception of /p/ was found to depend heavily on high frequency energy, but the differences in the perception of place for the two vowel environments suggest that the second formant transition contributed to perception with the vowel /I/. It has been shown that the perception of place in /d/ is enhanced by the presence of third formant transitions (Harris et al., 1958). The more acute perception of place for this consonant in the /u/ environment may also be the result of the absence of the third formant in the front vowel /I/.

TABLE 26

High cut-off frequencies of the second formant band-passes for the groups of consonants in the two vowel environments.

Consonant Group	Vowels /I/	Vowels /u/
p-b-m	2040	1480
t-d-n	2200	1800
k-g	2200	1700
f-v	2100	1600
s-z	2200	1700
ʃ-ʒ	2200	1800
θ-ð	2200	1700
l	2160	1600

The extent of transition may also explain the more acute perception of /f/ with /I/, and of /s/ and /z/ with /u/. As both /s/ and /z/ depend on high frequency invariant energy, a better performance would be expected with the vowel /I/ since the band-pass filter is including more invariant energy. The results suggest that the transition contributed to the perception of these consonants as was suggested by Delattre, Liberman and Cooper (1964). The extent of the transitions was also greater with /u/ for /θ/ and /ð/, therefore enhancing the perception of place, and these trends were also apparent in the unfiltered material.

The fricatives /ʃ/ and /ʒ/ had consistently fewer errors in the /I/ environment, which also produces smaller transitions than /u/. Both /ʃ/ and /ʒ/ are dependent on invariant energy and the improved performance with /I/ was the result of a greater inclusion of invariant energy (Table 26, p. 136) by the band-pass filter.

The results for manner and voicing had not been anticipated. In conditions containing the second formant, manner was best perceived for some consonants in the vowel environment which had also given the most acute perception of place and intelligibility. These results indicate that the perception of one feature influenced the perception of another (Pickett and Rubenstein, 1960). Manner was best discriminated for /m/ and /l/ with /I/ and can be attributed to the inclusion of the third formant resonance for /m/ (Potter, Kopp and Kopp, 1966) and to a more optimal transition rate for both /l/ and /m/ with /I/ (Cooper, et al., 1952; O'Connor et al., 1957).

When only the first formant was present, the smaller number of manner and voicing errors which occurred with the vowel /I/ for some fricatives and resonants were probably the result of the higher cut-off frequency of the band-pass filter for /I/ (Appendix A). The more acute voicing perception of /p/ and /b/ with /u/ cannot satisfactorily be explained. It is possible that the rate of first formant transition for /u/ was more optimal for the perception of voicing in these two consonants (Summerfield and Haggard, 1972).

Positions

In view of the findings of natural speech studies which indicated that transitions contribute most to the identification of place in final consonants (Beiter and Sharf, 1972; Fischer-Jorgensen, 1972a; Sharf and Hemeyer, 1972; Winitz, Scheib and Reeds, 1972), it had been hypothesized that differences in place errors and therefore intelligibility would occur between the two positions for those conditions containing the second formant. The differences encountered for the consonants in the two positions substantiated the hypothesis. Table 27 shows the positions which gave the most acute intelligibility and most acute perception of place for the consonants in the filter conditions containing the second formant. The upper cut-off frequencies of the second formant band-passes were also analyzed for the two positions. The results indicated that the cut-off frequencies were very similar for the two positions and therefore that the results obtained were not a consequence of the technique used.

The consonants /p, t, d, z/ showed consistently fewer place errors and a higher intelligibility in the final position for all three conditions (Table 27). These results agree with those of Sharf and Hemeyer (1972). /p/, /t/ and /z/ were seen to depend on invariant energy, most of which was removed by filtering. The improved discrimination in the final position indicates that the transition contributes more to the

TABLE 27

Positions giving the smallest number of place errors and highest intelligibility scores for the consonants in the three conditions containing the second formant.

	Condition		
	F12L	F2	F12H
p	Final*	Final	Final
b			
t	Final	Final	Final
d	Final	Final	Final
k			
g			Final
f			
v	Initial*		Initial
s			
z	Final	Final	Final
ʃ			Final
ʒ			Final
θ			
ð	Initial*	Initial	Initial
m			
n			
l			

* Trend also apparent in the unfiltered material.

N.B. Blanks indicate that the differences in scores for the consonants in the two positions were not significant.

identification of place in the final than in the initial position. The consonants /g/, /ʃ/ and /ʒ/ only showed a more acute place discrimination in the final position in Condition F12H. /v/ was more intelligible with fewer place errors in the initial position for Conditions F12L and F12H, while /ð/ showed this same trend for all three conditions. Both /v/ and /ð/ have been shown to depend on second formant transitions (Harris, 1954; 1958). It has been suggested that a more acute perception of the place feature in the initial position is related to the importance of invariant energy (House, Williams, Hecker and Kryter, 1965; Sharf and Hemeyer, 1972). As some of the invariant energy was preserved in the filtered utterances of the present study, the results with respect to position are not as clear as those obtained by Sharf and Hemeyer (1972). Sharf and Hemeyer eliminated the invariant energy and presented the transition and vocalic segments to listeners. It is interesting to note that the most acute perception of place in the final position occurred in the present study for those consonants which had most of their invariant energy eliminated by filtering. Sharf and Hemeyer explained their results on the basis of greater phonetic assimilation in forward coarticulation than in backward coarticulation. The span in the articulatory movement included in the CV transition is not as great as that included in the VC transition.

Differences in the perception of manner for the two positions were also found. Manner was best perceived in the initial position for /v, s, z, θ, ð/ in Condition F12L indicating the importance of invariant energy (House, Williams, Hecker and Kryter, 1965; Sharf and Hemeyer, 1972). In Condition F12L manner and intelligibility were more acute for nasals in the final position possibly indicating the importance of transitions, although the trend did not occur for the other conditions. In Condition F2, the voiced plosives were more intelligible in the final position in accordance with a general tendency for these sounds to be more accurately identified in that position.

In Condition F1, when differences occurred, intelligibility, and the perception of place and manner were most acute in the initial consonants, indicating the importance of invariant energy in this condition. The exceptions were the plosives /p/, /d/ and /b/, which could have depended more on the first formant transition than the remaining consonants.

Perceptual Importance of the Phonetic Features of Place, Manner and Voicing

The results of the multidimensional scaling procedure indicated that in all four filter conditions, the most important dimensions underlying consonant perception were those of manner and voicing. Condition F12L contained a dimension separating the plosives from the non-plosives and a dimension separating /m, n, l/ from the remaining consonants. Condition F2 contained a dimension separating the fricatives from the other consonants, and although the resonance feature did not appear as a dimension, the groupings of consonants indicated that manner for /m, n, l/ and for the plosives was well perceived. Figure 5 (p. 98), which shows the plotting of the stress values relative to the number of dimensions for each filter condition, indicates that although the most acute elbow in the curve for Condition F2 occurred at two dimensions, three dimensions could also have been chosen to interpret the data. Most probably the dimensions would have been similar to those obtained for Condition F12L. In Condition F1 and F12H, the dimension separating the plosives from the other consonants was also apparent. Voicing was the other prominent dimension, occurring in all the filter conditions. Although the overall number of voicing errors was significantly greater in Conditions F1 and F2 than in the other two filter conditions, the weakening of the dimension was only apparent in Condition F2 as seen by the spatial proximity of voiced and voiceless cognates. This is explained by the first formant frequencies which contained more voicing information than the second formant frequencies for some consonants, therefore systematically changing the patterns of confusions as seen in Table 22, (p. 94). In all filter conditions the place feature was only apparent in the distance between consonants, not being represented as a separate dimension, even though there were significantly fewer place errors in Conditions F12L and F2. Two possible considerations arise, the first being that although place errors were fewest in Conditions F12L and F2, they were still more numerous than in the unfiltered condition; and the second being that perceptually place is never a prominent feature. The results of Condition F12L, which most closely approximated normal speech, were compared to Shepard's (1972) spatial representation of Miller and Nicely's (1955) data for their wide band (200-6500 Hz) condition. Shepard pooled the Miller and Nicely data for the six conditions of noise (S/N: -18 dB, -12 dB, -6 dB, 0 dB, 6 dB, 12 dB), even though the different noise conditions affected the number of overall confusions and more particularly place confusions. He found that the increased number of confusions had little or no effect on the internal pattern of those confusions and that the dimensions of voicing and nasality (manner) accounted for most of the variance in the data. These results therefore also indicate that the place feature is not a perceptually salient feature. It has been shown in a study of short-term memory recall, that consonants are coded not as units, but as a set of distinctive features (Wickelgren, 1966). Peters (1963), using the multidimensional technique analysed the results of consonant similarity rating by subjects, and found that in general, phonemes were first sorted out according to manner, followed by voicing, and that for some trained individuals, place was also important. Peters (1967) also found that consonant pairs played forward or backward provided similar distance measures, indicating that the spectral characteristics of a sound contributed greatly to the degree of similarity.

The lack of prominence of the place feature in this and other studies may be attributed to several factors. If speech is performed by a series of binary decisions, no grouping of consonants into a binary place feature will succeed, since there are more places of articulation than manners of production and voicing possibilities (Klatt, 1968). Acoustically the place feature is the least redundant of the three features. This is seen in the results of the present study, which indicate that most of the place information is contained at the level of the second formant, while both first and second formants contained some voicing and manner information. Finally, in the English language, manner of production is used more often than any other articulatory criterion (Denes, 1963) to distinguish words from one another. Manner, therefore, seems to carry the greatest perceptual load, followed by voicing and then place, suggesting that as listeners we are least trained in making place distinctions.

Implications of the Results of the Present Study With Respect to Current Theories on Speech Perception

The rapidity with which we perceive speech has been attributed by some investigators to the restructuring or encoding of phonemes at the acoustic level (Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967; Liberman, 1970). Acoustically, the encoding is represented by the formant transitions, which have mostly been studied in relation to synthetic plosives (Cooper et al., 1952; Liberman et al., 1954; Delattre et al., 1955; Harris et al., 1958). Variation of the second formant transition leads to differences in aurally perceived distinctions among stops. Variation of the first formant changes the voicing characteristics of plosives (Liberman, Delattre and Cooper, 1958) and although the first formant transition has not been systematically studied with respect to manner distinctions, it contributes to a more accurate synthesis of the different classes of consonants (Delattre, 1958; 1962).

The increase in manner errors in the present study in the absence of the first formant for some consonants and particularly for the voiced plosives indicates that the first formant frequencies contain a certain amount of specific manner information. A comparison of Conditions F1 and F2 and of the spatial representation of the consonants in both these conditions (Figure 7, p. 103, Figure 8, p. 106) showed that the first and the second formant frequencies each contain important manner information for plosives. Manner identification in the remaining consonants and more specifically the fricatives is dependent on the invariant energy at the level of the second formant.

The first formant frequencies contain more voicing information than the second formant frequencies for the voiced plosives and some voiced fricatives because of the presence of the first formant transition, voice bar and low frequency invariant energy. Voicing, however, was still relatively well discriminated in the absence of the first formant, indicating that the invariant energy also contained important voicing cues.

Most of the place information was contained in the second formant frequencies. Since some of the consonants which showed the fewest number of place errors in Condition F2 were characterized by important invariant energy, this suggests that the amount of invariant energy contained in the band-pass filter was an important factor in determining the accuracy with which place was perceived. In addition, the importance of transitions as seen in other natural speech studies appeared to be influenced by vowel context and consonant position. Only the consonants /b/ and /d/ seemed to be cued mostly by second formant transitions.

As discussed in the preceding section, listeners process speech by making manner distinctions, followed by voicing and finally place distinctions. An encoded theory of speech perception would indicate that manner is encoded and therefore that the first formant transition plays an important role in the aurally perceived distinctions among different classes of consonants. The results of the present study, however, indicate that invariant energy is more important than transitions for the identification of manner. Similarly, invariant energy plays an important part in the accurate identification of voicing and place. Transitions do contain information which contributes to the identification of consonants, but it may be questioned if this is their primary function. The results of the present study seem to be in greater agreement with the results of

Cole and Scott (1973b), which showed that the invariant energy of consonants contains more information, than with the results of research at the Haskins Laboratories. Cole and Scott (1973c) and Scott and Cole (1973) have suggested that transitional cues serve primarily to preserve the temporal order of acoustic segments in rapid ongoing speech.

One objection to the present study is that the importance of the first two formants was studied in isolated syllables, which are less encoded than running speech (Liberman, 1970). However, the importance of transitions in synthetic speech has also been studied in syllables with steady-state vowels. Spectrographic analyses of natural running speech reveals wide variations in patterns of transitions for identical consonants, overlapping patterns for different consonants (Lehiste and Peterson, 1961; Ohman, 1966) and rarely occurring steady-state vowels (Shearme and Holmes, 1962). In addition the anticipation of a following vowel (Lehiste and Shockey, 1972a; 1972b) or of a following consonant (Kuehn and Moll, 1972) may result in a modification of a transitional pattern. Although transitions are more visually prominent and invariant energy less salient in running speech, very little research has been performed on the role of these transitions in the identification of consonants. It is not known if transitions and invariant energy carry less information for consonants in running speech than in nonsense syllables. Syllables are less intelligible than running speech which depends on important contextual cues for identification (Lieberman, 1963; Pickett and Pollack, 1963).

Implications of Findings for Hearing Impairment

The interpretation of synthetic speech studies, which have stated that the second formant transition is probably the most important carrier of linguistic information, have had an impact on research aimed at finding new ways of presenting speech to hearing-impaired listeners (Thomas and Sparks, 1971; Pfannebecker, 1972). Severely impaired listeners generally perceive low frequencies better than high frequencies which may not even be within their audible range. Listeners having a hearing loss make more consonant errors than vowel errors and also have more difficulty in distinguishing the place feature than manner and voicing features (Rosen, 1962, Cox, 1969; Pickett, Martin, Johnson, Smith, Daniel, Willis and Otis, 1972). It has been shown that deaf listeners use the same cues for identifying voicing, manner and place as do normal-hearing listeners (LaBenz, 1956; Cox, 1969; Pickett <u>et al</u>., 1972; Bennett, 1973).

The discrimination of synthetic second formant transitions has been studied in deaf listeners to gain insight in their poor place discrimination abilities (Martin, Osberger and Pickett, 1972; Danaher, Osberger and Pickett, 1973). The results of the present study suggest that the ability of deaf listeners to perceive various forms of invariant energy should also be studied. In addition, the present results indicate that a de-emphasis of the first formant in natural speech which has been attempted with deaf listeners (Thomas and Sparks, 1971; Pfannebecker, 1972) would in fact be eliminating very little information. The information contained in the first formant frequencies is very much poorer than that contained in the second formant frequencies. Severely deaf listeners are therefore limited by the information content of the sounds they can perceive. The second formant frequencies might provide much more information if they were presented in the audible range of deaf listeners. Blesser (1972) transformed the high frequency energy of speech to low frequency energy and vice versa. The speech spectrum was initially limited to a 200-3000 Hz band and then rotated around a centre frequency of 1600 Hz. According to Blesser, this kind of transformation did not obscure the information contained in the frequency spectrum, but rearranged it to create a new signal which had a one-to-one correspondence to the old signal. After some practise, listeners were able to communicate with one another using the spectrally transformed speech. It would be interesting to see if such a technique, which would bring the second and third formants into the first formant range, would improve the discrimination of profoundly deaf listeners.

CONCLUSIONS

The first and second formants when presented together at a comfortable listening level provide enough information for about 60% accuracy in the identification of consonants in CV and VC syllables. The features of voicing and manner are accurately perceived, but place distinctions are less accurate in the absence of higher formants. Not all consonants are affected to the same extent by the loss of frequency information above the second formant, as can be predicted from the different distributions of energy in the invariant part of the spectrum.

The consonants in CV and VC syllables can be identified with about 25% accuracy on the basis of first formant information alone, and elimination of the first formant decreases intelligibility by only about 6%. The first formant transition and low frequency invariant energy primarily serve to contribute voicing information for voiced plosives and fricatives. The first formant also contributes some manner information, particularly for plosives, but contains very little place information. Therefore, it can be concluded that the first formant does make a small, but definite, contribution to consonant intelligibility and although much of this information may be redundant, it would be of use under difficult listening conditions.

The second formant provides enough information for about 55% accuracy in the identification of consonants in CV and VC syllables, and intelligibility decreases by almost 40% when the second formant frequencies are eliminated. The second formant frequencies, therefore, contribute important information to consonant identification. Some place information is available from the second formant transition and associated invariant energy for the majority of consonants. The second formant frequencies serve to distinguish fricatives from the other classes of consonants, and also contribute to plosive, nasal and liquid distinctions. This manner information is contained in the invariant energy and perhaps in the duration of the second formant transition. Although some voicing confusions are apparent when only the second formant is heard, voicing distinctions are relatively well perceived in terms of invariant energy and durational characteristics. The first and second formants contain redundant voicing information, since the elimination of either formant did not drastically reduce the perception of this feature.

When both formants are amplified and presented at a high sound pressure level, consonants are still well identified, with a decrease in accuracy of only about 5% relative to the comfortable level of presentation. The high sound pressure level has the general effect of increasing place confusions. Voicing and manner distinctions remain as accurate as in the comfortable level condition, although some manner confusions do occur for the voiced fricatives.

In all filter conditions, intelligibility and the accuracy with which the place, manner and voicing features are perceived depends on information carried by transitions and invariant energy. The relative importance of each of these cues varies as a function of the vowel environment and the position of the consonant.

The present study was original in that consonant information in isolated first and second formants was systematically assessed in natural speech for the first time. The results of this investigation are in very close agreement with other filtered speech studies (Miller and Nicely, 1955), as well as natural speech studies which have assessed the relative importance of transitions and invariant energy for consonant perception (Malécot, 1956; 1958; Wang, 1959; Fischer-Jorgensen, 1972a; 1972b; 1972c; Cole and Scott, 1973a). These previous results, together with the present results, indicate the necessity for caution when generalizing from the conclusions of synthetic speech to natural speech.

In natural speech a listener hears both invariant and transitional cues. The present results indicate that both types of cues carry information, but transition cues may not be as crucial to consonant perception as suggested by the results of synthetic speech studies. Since the band-pass filters used in this study did not permit precise isolation of the formant patterns, it would be interesting to repeat the experiment using a tracking technique (Atal and Hanauer, 1971; Olive, 1971), which would permit the comparison of cues in isolated formant patterns for both syllables and running speech. Further investigators might also take into consideration the question of whether other spectral attributes such as overall spectrum shape (Fant, 1968) and envelope cues (Cole and Scott, 1973b) may be more basic to perception than band-limited cues contained in formant patterns.

REFERENCES

Abramson, A. S. & Lisker, L. Voice onset time in stop consonants: Acoustic analysis and synthesis. Proceedings of the Fifth International Congress on Acoustics, Liège (1965).

Ahmed, R. & Agrawal, S. S. Significant features in the perception of (Hindi) consonants. J. Acoust. Soc. Amer., 45, (1969), 758-763.

Ainsworth, W. A. Perception of stop consonants in synthetic CV syllables. Lang. Speech, 11, (1968a), 139-156.

Ainsworth, W. A. First formant transitions and the perception of synthetic semi-vowels. J. Acoust. Soc. Amer., 44, (1968b), 689-694.

Andresen, B. S. On the perception of unreleased voiceless plosives in English. Lang. Speech, 3, (1960), 109-119.

Atal, B. S. & Hanauer, S. L. Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Amer., 50, (1971), 637-655.

Barrs, J. Voiceless fricative identification as a function of duration modification. J. Acoust. Soc. Amer., 35, (1963), 784 (Abstract).

Beiter, R. & Sharf, D. J. Effect of coarticulation on perception: Identification of consonants from vowel transitions played forward and backward. Paper presented at A.S.H.A. Convention, San Francisco (1972).

Bennett, C. W. Discrimination of stop consonants by severely hearing-impaired children. Unpublished doctoral dissertation, McGill University (1973).

Blesser, B. Speech perception under conditions of spectral transformation: I Phonetic characteristics. J. Speech Hear. Res., 15, (1972), 5-41.

Borst, J. M. Use of spectrograms for speech analysis and synthesis. J. Audio Eng. Soc., 4, (1956), 14-23.

Broad, D. J. & Fertig, R. H. Formant-frequency trajectories in selected CVC syllable nuclei. J. Acoust. Soc. Amer., 47, (1970), 1572-1582.

Chang, S. H., Pihl, G. E. & Wiren, J. The intervalgram as a visual representation of speech sounds. J. Acoust. Soc. Amer., 23, (1951), 675-679.

Cole, R. A. & Scott, B. The phantom in the phoneme: Invariant cues for stop consonants. University of Waterloo Research Report in Psychology No. 45, (1973a), 1-25. (In press: Perception and Psychophys.)

Cole, R. A. & Scott, B. Toward a theory of speech perception. University of Waterloo Research Report in Psychology No. 46, (1973b). (Recently published in Psychol. Rev. 81, (1974), 348-374).

Cole, R. A. & Scott, B. Perception of temporal order in speech: The role of vowel transitions. Can. J. Psychol. (In press, 1973c)

Cole, R. A. & Scott, B. Acoustic invariance for stop consonants. J. Acoust. Soc. Amer., 51, (1972), 112 (Abstract).

Cooper, F. S. Instrumental methods for research in phonetics. Proceedings of the Fifth International Congress of Phonetic Sciences, Munster (1964).

Cooper, F. S. Speech synthesizers. Proceedings of the Fourth International Congress of Phonetic Sciences, Helsinki (1961). The Hague: Mouton and Co. (1962), pp. 3-13.

Cooper, F. S. Spectrum analysis. J. Acoust. Soc. Amer., 22, (1950), 761-762.

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

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

Cox, B. P. The identification of unfiltered and filtered consonant vowel stimuli by sensorineural hearing-impaired persons. Unpublished doctoral dissertation, University of Pittsburgh, (1969).

Danaher, E. M., Osberger, M. J. & Pickett, J. M. Discrimination of formant frequency transitions in synthetic vowels. J. Speech Hear. Res., 16, (1973), 439-451.

Danaher, E. M. & Pickett, J. M. Backward masking in sensorineural subjects and its effect on syllable discrimination. Paper presented at the A.S.H.A. Convention, Detroit (1973).

Delattre, P.C. Le jeu des transitions des formants et la perception des consonnes. Proceedings of the Fourth International Congress of Phonetic Sciences, Helsinki (1961). The Hague: Mouton and Co., (1962), pp. 407-417.

Delattre, P.C. Les indices acoustiques de la parole: Premier rapport. Phonetica, 2, (1958), 108-118; 226-251.

Delattre, P. C. Les attributs acoustiques de la nasalité vocalique et consonantique. Studia Linguistica, 8, (1954), 103-109.

Delattre, P. C., Liberman, A. M. & Cooper, F. S. Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. Studia Linguistica, 18, (1964), 104-121.

Delattre, P. C., Liberman, A. M. & Cooper, F. S. Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Amer., 27, (1955), 769-773.

Denes, P. B. On the statistics of spoken English. J. Acoust. Soc. Amer., 35, (1963), 892-904.

Denes, P. B. Effect of duration on the perception of voicing. J. Acoust. Soc. Amer., 27, (1955), 761-764.

Denes, P. B. & Pinson, E. N. The Speech Chain. Baltimore: Waverly Press, (1963).

Dixon, W. J. (Ed.) Biomedical Computer Programs. University of California Publications in Automatic Computation, No. 2. Berkeley: University of California Press, (1970).

Dunn, H. K. Methods of measuring vowel formant bandwidths. J. Acoust. Soc. Amer., 33 (1961), 1737-1746.

Egan, J. P. Articulation testing methods. Laryngoscope, 58, (1948), 955-991.

Egan, J. P. & Wiener, F. M. On the intelligibility of bands of speech in noise. J. Acoust. Soc. Amer., 18, (1946), 435-441.

Fairbanks, G. Test of phonemic differentiation: The Rhyme Test. J. Acoust. Soc. Amer., 30, (1958), 596-599.

Fant, G. Analysis and synthesis of speech processes. In Manual of Phonetics. B. Malmberg (Ed.), Amsterdam: North Holland Publishing Co., (1968), pp. 173-277.

Fant, G. Auditory patterns of speech. In Models for the Perception of Speech and Visual Form: Proceedings of a Symposium. W. Wathen-Dunn (Ed.), Cambridge, Mass.: M.I.T. Press, (1967a), pp. 111-125.

Fant, G. On the predictability of formant levels and spectrum envelopes from formant frequencies. In Readings in Acoustic Phonetics. I. Lehiste (Ed.), Cambridge, Mass.: M.I.T. Press, (1967b), pp. 44-120.

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

Fant, G. Sound spectrography. Proceedings of the Fourth International Congress of Phonetic Sciences, Helsinki, (1961). The Hague: Mouton and Co., (1962b), pp. 14-33.

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

Ferguson, G. A. Statistical Analysis in Psychology and Education. Third Edition. New York: McGraw-Hill, (1971).

Fischer-Jorgensen, E. Identification of unreleased Danish stop consonants. In Perceptual Studies of Danish Stop Consonants: Annual Report VI. Institute of Phonetics, University of Copenhagen, (1972a), pp. 89-103.

Fischer-Jorgensen, E. Tape cutting experiments with stop consonants. J. Acoust. Soc. Amer., 52, (1972b), 132 (Abstract).

Fischer-Jorgensen, E. Tape cutting experiments with Danish stop consonants in initial position. In Perceptual Studies of Danish Stop Consonants: Annual Report VI. Institute of Phonetics, University of Copenhagen, (1972c), pp. 104-168.

Fischer-Jorgensen, E. Identification of initial stop consonants played at double and half speed. In Perceptual Studies of Danish Stop Consonants: Annual Report VI. Institute of Phonetics, University of Copenhagen, (1972d), pp. 169-176.

Fischer-Jorgensen, E. New techniques in acoustic phonetics. In Psycholinguistics. S. Saporta (Ed.), New York: Holt, Rhinehart and Winston, Inc., (1961), pp. 112-142.

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

Fletcher, H. Speech and Hearing in Communication. Princeton, N.J.: Van Nostrand Co., (1929).

Franklin, B. The effect of a low-frequency band (240-480 Hz) of speech on consonant discrimination in hearing-impaired subjects. Paper presented at the A.S.H.A. Convention, San Francisco, (1972).

Franklin, B. The effect of a low frequency band of speech on consonant discrimination. Unpublished doctoral dissertation, City University of New York, (1969a).

Franklin, B. The effect on consonant discrimination of combining a low-frequency passband in one ear and a high-frequency passband in the other ear. J. Audit. Res., 9, (1969b), 365-378.

French, N. R. & Steinberg, J. C. Factors governing the intelligibility of speech sounds. J. Acoust. Soc. Amer., 19 (1947), 90-119.

Fujimura, O. Analysis of nasal consonants. J. Acoust. Soc. Amer., 34, (1962), 1865-1875.

Gay, T. Effects of filtering and vowel environment on consonant perception. J. Acoust. Soc. Amer., 48, (1970), 993-998.

Gerstman, L. J. Noise duration as a cue for distinguishing among fricative, affricative and stop consonants. J. Acoust. Soc. Amer., 28, (1956), 160 (Abstract).

Greenhouse, S. W. & Geisser, S. On methods in the analysis of profile data. Psychometrika, 24, (1959), 95-112.

Grimm, W. A. Perception of segments of English spoken consonant vowel syllables. J. Acoust. Soc. Amer., 40, (1966), 1454-1461.

Gupta, J. P., Agrawal, S. S. & Ahmed, R. Perception of (Hindi) consonants in clipped speech. J. Acoust. Soc. Amer., 45, (1969), 770-773.

Haggard, M. The perception of speech. Report on research in progress in the Department of Psychology, Queen's University of Belfast, Northern Ireland, Series 2, No. 1, (1972).

Haggard, M., Ambler, S. & Callow, M. Pitch as a voicing cue. J. Acoust. Soc. Amer., 47, (1970), 613-617.

Haggard, M. & Summerfield Q. Perception of voicing - A supplementary report. Report on research in progress in the Department of Psychology, Queen's University of Belfast, Northern Ireland, Series 2, No. 1, (1972), pp. 37-38.

Halle, M., Hughes, G. W. & Radley, J. P. A. Acoustic properties of stop consonants. J. Acoust. Soc. Amer., 29, (1957), 107-116.

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

Harris, K. Some acoustic cues for the fricative consonants. J. Acoust. Soc. Amer., 28, (1956), 160 (Abstract).

Harris, K. Cues for the identification of the fricatives of American English. J. Acoust. Soc. Amer., 26, (1954), 952 (Abstract).

Harris, K., Hoffman, H. S., Liberman, A. M., Delattre, P. C. & Cooper, F. S. Effect of third formant transitions on the perception of the voiced stop consonants. J. Acoust. Soc. Amer., 30, (1958), 122-126.

Hayashi, R., Ohta, F. & Morimoto, M. Binaural fusion test: A diagnostic approach to the central auditory disorders. Internat. Audiol., V, (1966), 133-135.

Hecker, M. H. Studies of nasal consonants with an articulatory speech synthesizer. J. Acoust. Soc. Amer., 34, (1962), 179-188.

Heinz, J. M. & Stevens, K. N. On the properties of voiceless fricative consonants. J. Acoust. Soc. Amer., 33, (1961), 589-596.

Hirsh, I. J., Reynolds, E. G. & Joseph, M. Intelligibility of different speech materials. J. Acoust. Soc. Amer., 26, (1954), 530-538.

Hoffman, H. S. Study of some cues in the perception of the voiced stop consonant. J. Acoust. Soc. Amer., 30, (1958), 1035-1041.

House, A. S. & Fairbanks, G. The influence of consonant environment upon secondary acoustical characteristics of vowels. J. Acoust. Soc. Amer., 25, (1953), 128-136.

House, A. S., Williams, C. E., Hecker, M. H. L. & Kryter, K. D. Articulation testing methods: Consonantal differentiation with a closed-response set. J. Acoust. Soc. Amer., 37, (1965), 158-166.

House, H. Analog studies of nasal consonants. J. Speech Dis., 22, (1957), 190-204.

Hughes, G. W. & Halle, M. Spectral properties of fricative consonants. J. Acoust. Soc. Amer., 28, (1956), 303-310.

Jakobson, R. & Halle, M. Fundamentals of Language. The Hague: Mouton and Co., (1956).

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

Klahr, D. A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure. Psychometrika, 34, (1969), 319-330.

Klatt, D. H. Structure of confusions in short-term memory between English consonants. J. Acoust. Soc. Amer., 44, (1968), 401-407.

Koenig, W. H., Dunn, H. K. & Lacey L. Y. The sound spectrograph. J. Acoust. Soc., 18, (1946), 19-49.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, (1964a), 1-27.

Kruskal, J. B. Nonmetric multidimensional scaling: a numerical method. Psychometrika, 29, (1964b), 115-129.

Kryter, K. D. Speech bandwidth compression through spectrum selection. J. Acoust. Soc. Amer., 32, (1960), 547-556.

Kuehn, D. P. & Moll, K. L. Perceptual effects of forward coarticulation. J. Speech Hear. Res., 15, (1972), 654-664.

LaBenz, P. J. Potentialities of auditory perception for various levels of hearing loss. Volta Rev., 58, (1956), 397-402.

Latif, A., Gallagher, T., Goldstein, J. & Daniloff, R. Perception of coarticulated nasality. J. Acoust. Soc. Amer., 49, (1971), 538-540.

Lehiste, I. & Peterson, G. E. Transitions, glides and diphthongs. J. Acoust. Soc. Amer., 33, (1961), 268-277.

Lehiste, I. & Shockey, L. Coarticulation effects in the identification of final plosives. J. Acoust. Soc. Amer., 51, (1972a), 101 (Abstract).

Lehiste, I. & Shockey, L. On the perception of coarticulation effects in English VCV syllables. J. Speech Hear. Res., 15, (1972b), 500-506.

Liberman, A. M. Some characteristics of perception in the speech mode. In Perception and Its Disorders Research Publication, XLVIII, Association for Research in Nervous and Mental Disease, (1970), 238-254.

Liberman, A. Some results of research on speech perception. J. Acoust. Soc. Amer., 29, (1957), 117-123.

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

Liberman, A. M., Delattre, P. C. & Cooper, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. Lang. Speech, 1, (1958), 153-167.

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

Liberman, A. M., Delattre, P. C., Cooper, F. S. & Gerstman, L. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychol. Monogr., 68, (1954), 1-16.

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

Liberman, A. M., Harris, K., Hoffman, H. & Griffith, B. The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol., 54, (1957), 358-368.

Liberman, A. M., Ingemann, F., Lisker, L., Delattre, P. & Cooper, F. S. Minimal rules for synthesizing speech. J. Acoust. Soc. Amer., 31, (1959), 1490-1499.

Licklider, J. C. R. The effects of amplitude distortion upon the intelligibility of speech. J. Acoust. Soc. Amer., 18, (1946), 429-434.

Licklider, J. C. R. & Miller, G. A. The perception of speech. In Handbook of Experimental Psychology. S. S. Stevens (Ed.) New York: Wiley & Sons, (1960), pp. 1040-1074.

Licklider, J. C. R. & Pollack, I. Effects of differentiation, integration and infinite peak clipping upon the intelligibility of speech. J. Acoust. Soc. Amer., 20, (1948), 42-51.

Lieberman, P. On the redundancy of speech and articulation and perception. J. Acoust. Soc. Amer., 35, (1963), 807 (Abstract).

Lindblom, B. Accuracy and limitations of sonograph measurements. Proceedings of the Fourth International Congress of Phonetic Sciences, Helsinki (1961). The Hague: Mouton and Co., (1962), pp. 188-202.

Linden, A. Distorted speech and binaural speech resynthesis tests. Acta-Otol. (Stockh.), 58, (1964), 32-47.

Lintz, L. B. & Sherman, D. Phonetic elements and perception of nasality. J. Speech Hear. Res., 4, (1961), 381-396.

Lisker, L. Minimal cues for separation of /w, r, l, j/ in intervocalic position. Word, 13, (1957), 256-267.

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

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

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

Lisker, L. & Abramson, A. S. Stop categorization and voice onset time. Proceedings of the Fifth International Congress of Phonetic Sciences, Munster, (1964b).

Malécot, A. The role of releases in the identification of released final stops. Language, 34, (1958), 370-380.

Malécot, A. Acoustic cues for nasal consonants. Language, 32, (1956), 274-284.

Martin, E., Osberger, M. J. & Pickett, J. M. Discrimination loss and masking by low frequency vowel formants. Paper presented at the Meeting of the Acoustical Society of America, Buffalo, N.Y., (1972).

Martin, E. S., Pickett, J. M. & Colten, S. Discrimination of vowel formant transitions by listeners with severe sensorineural hearing loss. In Speech Communication Ability and Profound Deafness. G. Fant (Ed.), Washington, D.C.: A. G. Bell Assoc., (1972), pp. 81-98.

Matzker, J. The binaural test. Internat. Audiol., 1, (1962), 209-211.

Miller, G. A. & Nicely, P. E. An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Amer., 27, (1955), 338-352.

Morgan, B. J. T. Cluster analyses of two acoustic confusion matrices. Perception and Psychophys., 13, (1973), 13-24.

Nakata, K. Synthesis and perception of nasal consonants. J. Acoust. Soc. Amer., 31, (1959), 661-666.

O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C. & Cooper, F. S. Acoustic cues for the perception of initial /w, j, r, l/ in English. Word, 13, (1957), 25-43.

Ohman, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. J. Acoust. Soc. Amer., 39, (1966), 151-168.

Olive, J. Automatic formant tracking by a Newton-Raphson technique. J. Acoust. Soc. Amer., 50, (1971), 661-670.

Overall, J. Note on the scientific status of factors. Psychol. Bull., 61, (1964), 270-276.

Palva, A. Filtered speech audiometry. Basic studies with Finnish speech towards the creation of a method for the diagnosis of central hearing disorders. Acta-Otol. Suppl. 210, (1965), 1-86.

Peters, R. Perceptual organization for speech and other auditory signals. Aerospace Medical Research Laboratories, TR-68-31, Wright-Patterson Air Force Base, Ohio, (1967).

Peters, R. Dimensions of perception for consonants. J. Acoust. Soc. Amer., 35, (1963), 1985-1989.

Peterson, G. E. & Barney, H. L. Control methods used in a study of the vowels. J. Acoust. Soc. Amer., 24, (1952), 175-184.

Peterson, G. E. & Lehiste, I. Duration of syllable nuclei in English. J. Acoust. Soc. Amer., 32, (1960), 693-703.

Pfannebecker, G. B. Discrimination of spectrally weighted speech by hearing-impaired subjects. Unpublished M.Sc. thesis, University of Massachussetts, (1972).

Pickett, J. M. Sound patterns of speech: An introductory sketch. Amer. Ann. Deaf., 113, (1968), 120-133.

Pickett, J. M. Some acoustic cues for synthesis of the /n-d/ distinction. J. Acoust. Soc. Amer., 38, (1965), 474-477.

Pickett, J. M. & Danaher, E. M. On discrimination of formant transitions by persons with severe sensorineural hearing loss. Proceedings of the Symposium on Auditory Analysis and Perception of Speech, Leningrad, (1973).

Pickett, J. M., Martin, E. S., Johnson, D., Smith, S. B., Daniel, Z., Willis, D. & Otis, W. On patterns of speech feature reception by deaf listeners. In Speech Communication Ability and Profound Deafness. G. Fant (Ed.), Washington, D.C.: A. G. Bell Assoc., (1972), pp. 119-133.

Pickett, J. M. & Pollack, I. Adjacent context and the intelligibility of words excised from fluent speech. J. Acoust. Soc. Amer., 35, (1963), 807 (Abstract).

Pickett, J. M. & Rubenstein, H. Perception of consonant voicing in noise. Lang. Speech, 3, (1960), 155-163.

Potter, R. K., Kopp, G. A. & Green, H. C. Visible Speech. New York: Van Nostrand, (1947).

Potter, R. K., Kopp, G. A. & Kopp, H. G. Visible Speech. New York: Dover Publications, (1966).

Raphael, L. J. Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. J. Acoust. Soc. Amer., 51, (1972), 1296-1303.

Rosen, J. Phoneme identification in sensorineural deafness. Unpublished doctoral dissertation, Stanford University, (1962).

Schatz, C. D. The role of context in the perception of stops. Language, 30, (1954), 47-56.

Scheib, M. E. & Winitz, H. Identification of stops and vowels for the burst portion of /p, t, k/. J. Acoust. Soc. Amer., 51, (1972), 102 (Abstract).

Scott, B. & Cole. R. A. Satiation of consonant features in consonant-vowel syllables. University of Waterloo Research Report in Psychology No. 45, (1973), 41-59. (In press: Percept. & Psychophys.)

Sharf, D. J. Perceptual parameters of consonant sounds. Lang. Speech, 14, (1971), 169-177.

Sharf, D. J. Distinctiveness of 'defective' fricative sounds. Lang. Speech, 11, (1968), 38-45.

Sharf, D. J. & Hemeyer, T. Identification of place of consonant articulation from vowel formant transitions. J. Acoust. Soc. Amer., 51, (1972), 652-658.

Shearme, J. N. & Holmes, J. N. An experimental study of the classification of sounds in continuous speech according to their distribution in the Formant 1 - Formant 2 plane. Proceedings of the Fourth International Congress of Phonetic Sciences, Helsinki, (1961). The Hague: Mouton and Co., (1962), pp. 232-240.

Shepard, R. N. Psychological representation of speech sounds. In Human Communication: A Unified View. E. E. David & P. B. Denes (Eds.), New York: McGraw-Hall, (1972), pp. 67-113.

Singh, S. Cross-language study of perceptual confusions of plosive phonemes in two conditions of distortion. J. Acoust. Soc. Amer., 40, (1966), 635-656.

Singh, S. & Woods, D. R. Perceptual structure of 12 American English vowels. J. Acoust. Soc. Amer., 49, (1971), 1861-1866.

Stevens, K. N. & House, A. S. Studies of formant transitions using vocal tract analog. J. Acoust. Soc. Amer., 28, (1956), 578-585.

Stevens, K. N., House, A. S. & Paul, A. P. Acoustical description of syllabic nuclei: An interpretation interms of a dynamic model of articulation. J. Acoust. Soc. Amer., 40, (1966), 123-132.

Stevens, K. N., Kasowski, S. & Fant, G. Electrical analog of the vocal tract. J. Acoust. Soc. Amer., 25, (1953), 734-742.

Stevens, K. N. & Klatt, D. H. Role of formant transitions in the voiced-voiceless distinction of stops. J. Acoust. Soc. Amer., 55, (1974), 653-659.

Stevens, K. N. & Klatt, D.H. The role of formant transitions in the voice-voiceless distinction for stops. M.I.T. Quart. Prog. Report No. 101 (1971), 188-197.

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

Summerfield, Q. & Haggard, M. P. Perception of stop voicing - A rate-specific transition detector. Report on research in progress in the Department of Psychology, Queen's University of Belfast, Northern Ireland, Series 2, No. 1, (1972), 1-14.

Thomas, I. The influence of first and second formants on the intelligibility of clipped speech. J. Audio Eng. Soc., 16, (1968a).

Thomas, I. B. Single formant synthetic speech. I.E.E.E. Transactions on Audio and Electroacoustics, AU-16, (1968b), 288.

Thomas, I. B. The significance of the second formant in speech intelligibility. Biological Computer Laboratory, University of Illinois, Technical Report No. 10, (1966).

Thomas, I. B. & Niederjohn, R. J. The intelligibility of filtered-clipped speech in noise. J. Audio Eng. Soc., 18, (1970), 299-303.

Thomas, I. B. & Niederjohn, R. J. Enhancement of speech intelligibility at high noise levels by filtering and clipping. J. Audio Eng. Soc., 16, (1968), 412-415.

Thomas, I. B. & Ohley, W. J. Intelligibility enhancement through spectral weighting. Proceedings of the International Conference on Speech Communication and Processing, Newton, Mass., (1972).

Thomas, I. B. & Ravindran, A. Preprocessing of an already noisy speech signal for intelligibility enhancement. J. Acoust. Soc. Amer., 49, (1971), 133 (Abstract).

Thomas, I. B. & Sparks, D. W. Discrimination of filtered clipped speech by hearing-impaired subjects. J. Acoust. Soc. Amer., 49, (1971), 1881-1888.

Thomas, I. B. & Sparks, D. W. Discrimination of filtered clipped speech by sensorineural hearing-impaired subjects. J. Acoust. Soc. Amer., 47, (1970), 128 (Abstract).

Wang, W. S. Y. Transition and release as perceptual cues for final plosives. J. Speech Hear. Res., 2, (1959), 66-73.

Wang, W. S. Y. & Fillmore C. J. Intrinsic cues and consonant perception. J. Speech Hear. Res., 4, (1961), 130-136.

Wickelgren, W. A. Distinctive features and errors in short-term memory for English consonants. J. Acoust. Soc. Amer., 39, (1966), 388-398.

Winer, B. J. (Ed.) Statistical Principles in Experimental Design. Second Edition. New York: McGraw-Hill, (1971).

Winitz, H., Scheib, M. E. & Reeds, J. A. Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. J. Acoust. Soc. Amer., 51, (1972), 1309-1317.

APPENDIX A

RESULTS OF THE SPECTROGRAPHIC ANALYSIS FOR THE 136 SYLLABLES

TABLES 1-8

TABLE 1

List 1 (CV/I/JD)

Consonant	High F1 (Hz)	Invariant F2 (Hz)	Vocalic: Low F2 (Hz)	Vocalic: High F2 (Hz)	Low F3 (Hz)	Band-Pass Cut-Off Frequencies (Hz)
p	520	1020 R	2000	2040	2760	F1: 140-520
b	560	1020 R	1800	2000	2640	F2: 900-1950
m	560	870 R	1800	2000	2500	
t	520	1920	1920	2000	2760	F1: 140-560
d	480	1720	1840	2000	2640	F2: 1720-2000
n	560	1800	1920	2040	2600	
k	440	2500	1800	2160	2680	F1: 140-600
g	600	2500	2000	2200	2720	F2: 1800-2100
f	600	1020	1960	1960	2680	F1: 140-600
v	600	1600	1880	1880	2600	F2: 1020-1960
s	600		1800	2040	2800	F1: 140-600
z	560	1600	1600	1880	2800	F2: 1600-2040
ʃ	600	2000	1800	2000	2520	F1: 150-600
ʒ	600	2000	1640	1640	2720	F2: 1640-1900
θ	600		1840	2000	2800	F1: 140-600
ð	600	1520	1600	1920	2680	F2: 1520-2000
l	520	1320	1520	1920	2800	F1: 140-520 F2: 1320-1920

R = Reported value (Potter, Kopp and Kopp, 1966).

TABLE 2

List 2 (VC/I/JD)

Consonant	High F1 (Hz)	Invariant F2 (Hz)	Vocalic: Low F2 (Hz)	Vocalic: High F2 (Hz)	Vocalic: Low F3 (Hz)	Band-Pass Cut-Off Frequencies (Hz)
p	560		1680	2240	2600	F1: 140-640
b	640	1000	1520	2360	2440	F2: 1000-1900
m	560	1160	1480	2200	2520	
t	640		1600	2200	2920	F1: 130-640
d	520	1880	1720	2200	2960	F2: 1400-2200
n	520	1400	1600	2200	3000	
k	520	2000	20[illegible]	2240	2400	F1: 130-520
g	320	2320	2000	2240	2800	F2: 1700-1900
f	560	1000	1520	2080	2600	F1: 130-560
v	440	1320	1440	2120	2400	F2: 1000-1800
s	520	1720 R	1640	2080	2800	F1: 130-520
z	440	1720 R	1600	2200	2800	F2: 1600-2200
ʃ	520	2100	1920	2160	2800	F1: 130-520
ʒ	400	2160	1800	2100	2800	F2: 1800-2160
θ	600		1680	2200	2720	F1: 130-600
ð	400	1440	1680	2200	2680	F2: 1440-2100
l	480	920	2120	2120	2600	F1: 130-480 F2: 920-2000

TABLE 3

List 3 (CV/u/JD)

Consonant	High F1 (Hz)	Invariant F2 (Hz)	Vocalic: Low F2 (Hz)	Vocalic: High F2 (Hz)	Low F3 (Hz)	Band-Pass Cut-Off Frequencies (Hz)
p	400		1080	1200	2120	
b	480	1000	1040	1200	2040	F1: 140-480
m	420	800	1080	1320	2200	F2: 800-1320
t	360	1920	960	1640	2160	
d	400		1000	1760	2200	F1: 130-400
n	400	1740	880	1600	2120	F2: 880-1700
k	400	1400	840	1280	2200	F1: 140-400
g	400	1400	1000	1000	2240	F2: 840-1400
f	400		800	1280	2200	F1: 140-400
v	400	1320	1000	1400	2200	F2: 800-1400
s	400		1000	1680	2280	F1: 130-400
z	400	1680	1080	1680	2200	F2: 1000-1680
ʃ	440	2000	1000	1800	2200	F1: 150-440
ʒ	440	2000	1000	1800	2200	F2: 1000-1700
θ	400		1000	1640	2200	F1: 140-400
ð	360	1520	1000	1520	2200	F2: 1000-1640
l	480	1480	1000	1600	2280	F1: 150-480 F2: 1000-1600

TABLE 4

List 4 (VC/u/JD)

Consonant	High F1 (Hz)	Invariant F2 (Hz)	Vocalic Low F2 (Hz)	Vocalic High F2 (Hz)	Vocalic Low F3 (Hz)	Band-Pass Cut-Off Frequencies (Hz)
p	400		920	1040	2200	
b	400	1480	1000	1160	2200	F1: 140-400
m	400	800	1080	1080	2200	F2: 800-1480
t	400		880	1200	2280	
d	400	1880	1080	1240	2280	F1: 150-480
n	480		1160	1240	2280	F2: 880-1800
k	400	1520	1000	1120	2200	F1: 150-400
g	400	1840	960	1120	2200	F2: 960-1700
f	560		1080	1200	2280	F1: 150-560
v	480	1480	1040	1160	2200	F2: 1040-1480
s	440		1120	1240	2200	F1: 160-440
z	400	1800	1080	1280	2200	F2: 1080-1700
ʃ	400	2000	1120	1240	2280	F1: 160-400
ʒ	400	2000	960	1040	2280	F2: 960-1800
θ	480	1600 R	1120	1800	2240	F1: 150-480
ð	400	1600 R	960	1080	2200	F2: 960-1700
l	400	880	880	960	2200	F1: 150-400 F2: 800-1000

TABLE 5

List 5 (CV/I/RS)

Consonant	High F1 (Hz)	Invariant F2 (Hz)	Vocalic: Low F2 (Hz)	Vocalic: High F2 (Hz)	Low F3 (Hz)	Band-Pass Cut-Off Frequencies (Hz)
p	600	1760	1840	1960	2640	
b	520	1400	1800	2000	2200	F1: 170-600
m	600	1000	1680	2000	2320	F2: 1000-1700
t	600	1720 R	1960	2000	2760	
d	600	1720 R	1840	2040	2800	F1: 170-600
n	600	1720 R	1840	2200	2840	F2: 1720-2200
k	600	2320	2080	2160	2800	F1: 170-600
g	600	2480	1920	2200	2880	F2: 1920-2200
f	680	1020 R	1720	2000	2640	F1: 170-640
v	600	1020 R	2080	2080	2200	F2: 1050-1720
s	640	1520	2120	2120	2880	F1: 140-640
z	640	1600	1720	2160	2960	F2: 1520-2160
ʃ	600	2440	2040	2040	2800	F1: 150-600
ʒ	600	2280	2000	2000	2800	F2: 2000-2200
θ	600	1600 R	1760	2200	2800	F1: 140-680
ð	680	1600 R	2200	2200	2760	F2: 1600-2100
l	680	1720	2160	2160	2880	F1: 140-680 F2: 1720-2160

TABLE 6

List 6 (VC/I/RS)

Consonant	High F1 (Hz)	Invariant F2 (Hz)	Vocalic: Low F2 (Hz)	Vocalic: High F2 (Hz)	Low F3 (Hz)	Band-Pass Cut-Off Frequencies (Hz)
p	600	1000	1520	2120	2400	F1: 150-600
b	600	1440	2160	2160	2600	F2: 1000-1900
m	600	1360	2200	2200	2520	
t	600		2000	2160	2880	F1: 160-600
d	600	1800	2080	2080	2800	F2: 1700-2100
n	600	1700	2040	2100	2720	
k	520	2000	2120	2240	2600	F1: 160-520
g	520	2480	2240	2240	2680	F2: 1800-2000
f	600		1800	2160	2800	F1: 160-600
v	600	1800	2200	2200	2720	F2: 1800-2100
s	640	1600	2080	2200	2920	F1: 160-640
z	520	1640	2200	2200	2840	F2: 1600-2200
ʃ	600	2040	2040	2280	2760	F1: 170-600
ʒ	520		2320	2320	2800	F2: 1900-2100
θ	640	1600	2080	2200	2920	F1: 160-640
ð	520	1640	2200	2200	2840	F2: 1600-2200
l	600	1120	2080	2080	2800	F1: 170-600 F2: 1120-2080

TABLE 7

List 7 (CV/u/RS)

Consonant	High F1 (Hz)	Invariant F2 (Hz)	Vocalic Low F2 (Hz)	Vocalic High F2 (Hz)	Low F3 (Hz)	Band-Pass Cut-Off Frequencies (Hz)
p	480		1200	1400	2520	
b	440	1000	1200	1360	2280	F1: 160-480
m	440	900	1320	1320	2200	F2: 900-1400
t	480	1760	1320	1320	2280	
d	480	1720	1320	1320	2400	F1: 180-480
n	480	1520	1280	1640	2400	F2: 1280-1760
k	480	1680	1000	1400	2240	F1: 170-560
g	560	1680	1320	1320	2000	F2: 1000-1500
f	580		1020	1400	2400	F1: 180-580
v	560	1120	1400	1400	2400	F2: 1020-1400
s	480		1500	1640	2520	F1: 180-480
z	480	1320	1520	1700	2600	F2: 1320-1700
ʃ	560	2000	1320	1320	2200	F1: 180-560
ʒ	520	2000	1200	1200	2400	F2: 1200-1700
θ	520	1200	1120	1600	2400	F1: 160-520
ð	480		1280	1480	2480	F2: 1120-1600
l	480	1040	1320	1480	2400	F1: 180-480 F2: 1040-1480

TABLE 8

List 8 (VC/u/RS)

Consonant	High F1 (Hz)	Invariant F2 (Hz)	Vocalic: Low F2 (Hz)	Vocalic: High F2 (Hz)	Low F3 (Hz)	Band-Pass Cut-Off Frequencies (Hz)
p	520		880	1000	2200	
b	480	840	920	1000	2200	F1: 160-520
m	480	800	1000	1000	2200	F2: 800-1000
t	520	1800	1120	1520	2480	
d	480	1600	1000	1000	2200	F1: 180-520
n	440		1120	1320	2200	F2: 1000-1700
k	400	1200	800	960	2200	F1: 180-440
g	440	1800	1000	1200	2200	F2: 800-1700
f	440	1600	1080	1200	2200	F1: 170-440
v	440	1520	1200	1360	2200	F2: 1080-1600
s	440	1640	920	1400	2280	F1: 160-480
z	480	1520	1200	1200	2200	F2: 920-1640
ʃ	440	2000	1200	1200	2180	F1: 160-480
ʒ	480	2000	1000	1320	2200	F2: 1000-1700
θ	480	1800	1040	1400	2200	F1: 160-480
ð	480	1440	1200	1200	2200	F2: 1040-1700
l	440	900	920	1000	2400	F1: 190-440 F2: 900-1100

APPENDIX B

MINIMUM ATTENUATION OF UNWANTED FORMANTS BY SECOND FORMANT BAND-PASS

TABLES 1-8

TABLE 1

List 1 (CV/I/JD)

		ATTENUATION (dB)*				
				Vocalic		
Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	High F1	Invariant F2	Low F2	High F2	Low F3
p-b-m	900-1950	50	6	0**	10	30
t-d-n	1720-2000	75+	3	0	6	30
k-g	1800-2100	75+	20	3	9	30
f-v	1020-1960	57	3	3	3	33
s-z	1600-2040	75+	3	3	3	43
ʃ-ʒ	1640-1900	75+	8	3	8	30
θ-ð	1520-2000	75+	3	0	3	36
l	1320-1920	75+	3	0	3	46

* Relative to centre frequency of second formant band-pass.

** Attenuation less than 3 dB (cut-off frequencies) listed as 0dB.

TABLE 2

List 2 (VC/I/RS)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB) High F1	Invariant F2	Vocalic Low F2	Vocalic High F2	Low F3
p-b-m	1000-1900	50	3	0	20	30
t-d-n	1400-2200	75+	3	0	3	36
k-g	1700-1900	75+	22	11	21	30
f-v	1000-1800	64	3	0	20	30
s-z	1600-2200	75+	0	3	3	30
ʃ-ʒ	1800-2160	75+	3	3	3	36
θ-ð	1440-2100	75+	3	0	10	30
l	920-2000	70	3	10	10	30

TABLE 3

List 3 (CV/u/JD)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB)				
				Vocalic		
		High F1	Invariant F2	Low F2	High F2	Low F3
p-b-m	800-1320	57	3	0	3	50
t-d-n	880-1700	75+	14	5	7	30
k-g	840-1400	75+	3	3	0	50
f-v	800-1400	75	0	3	3	50
s-z	1000-1680	75+	3	3	3	30
ʃ-ʒ	1000-1700	75+	18	3	6	30
θ-ð	1000-1640	75+	0	3	3	36
l	1000-1600	75+	0	3	3	38

TABLE 4

List 4 (VC/u/JD)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB)				
				Vocalic		
		High F1	Invariant F2	Low F2	High F2	Low F3
p-b-m	800-1480	75	3	0	0	43
t-d-n	880-1800	70	8	3	0	30
k-g	960-1700	75+	12	3	0	30
f-v	1040-1480	66	3	3	0	46
s-z	1080-1700	75+	8	3	0	30
ʃ-ʒ	960-1800	75+	14	3	0	30
θ-ð	960-1700	75	0	3	8	30
l	800-1000	75	0	0	0	75+

TABLE 5

List 5 (CV/I/RS)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB)				
				Vocalic		
		High F1	Invariant F2	Low F2	High F2	Low F3
p-b-m	1000-1700	56	6	10	18	30
t-d-n	1720-2200	75+	3	0	3	32
k-g	1920-2200	75+	17	3	3	30
f-v	1050-1720	54	6	23	23	30
s-z	1520-2160	75+	3	0	3	30
ʃ-ʒ	2000-2200	75+	14	3	3	30
θ-ð	1600-2100	75+	3	8	8	30
l	1720-2160	75+	3	3	3	30

TABLE 6

List 6 (VC/I/RS)

		ATTENUATION (dB)				
				Vocalic		
Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	High F1	Invariant F2	Low F2	High F2	Low F3
p-b-m	1000-1900	58	3	18	18	30
t-d-n	1700-2100	75+	3	0	6	30
k-g	1800-2000	75+	24	15	15	30
f-v	1800-2100	75+	3	6	6	30
s-z	1600-2200	75+	3	3	3	30
ʃ-ʒ	1900-2100	75+	0	14	14	30
θ-ð	1600-2200	75+	3	3	3	30
l	1120-2080	68	3	3	3	36

TABLE 7

List 7 (CV/u/RS)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB) High F1	Invariant F2	Vocalic Low F2	Vocalic High F2	Low F3
p-b-m/	900-1400	68	3	0	3	50
t-d-n	1280-1760	75+	3	3	0	30
k-g	1000-1500	66	13	3	0	30
f-v	1020-1400	63	0	3	3	58
s-z	1320-1700	75+	3	0	3	46
ʃ-ʒ	1200-1700	75+	21	3	3	30
θ-ð	1120-1600	75+	0	3	3	46
l	1040-1480	75+	3	0	3	53

TABLE 8

List 8 (VC/u/RS)

		ATTENUATION (dB)				
				Vocalic		
Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	High F1	Invariant F2	Low F2	High F2	Low F3
p-b-m	800-1000	53	3	3	3	75+
t-d-n	1000-1700	73	8	3	3	30
k-g	800-1700	72	7	3	0	30
f-v	1080-1600	75+	3	3	0	38
s-z	920-1640	71	3	3	0	33
ʃ-ʒ	1000-1700	75+	20	3	0	30
θ-ð	1040-1700	75+	8	3	0	30
l	900-1100	75+	3	0	0	75+

APPENDIX C

MINIMUM ATTENUATION OF UNWANTED FORMANTS BY FIRST FORMANT BAND-PASS

TABLES 1-8

TABLE 1

List 1 (CV/I/JD)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB)* High F1	Low F2
p - b - m	140-520	10	50
t - d - n	140-560	3	75+
k - g	140-600	3	75+
f - v	140-600	3	60
s - z	140-600	3	75+
ʃ - ʒ	150-600	3	75+
θ - ð	140-600	3	75+
l	140-520	3	75+

* Relative to centre frequency of first formant band-pass.

TABLE 2

List 2 (VC/I/JD)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB)	
		High F1	Low F2
p - b - m	140-640	3	54
t - d - n	130-640	3	75+
k - g	130-520	3	75+
f - v	130-560	3	64
s - z	130-520	3	75+
ʃ - ʒ	130-520	3	75+
θ - ð	130-600	3	75+
l	130-480	3	73

TABLE 3

List 3 (CV/u/ðD)

		ATTENUATION (dB)	
Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	High F1	Low F2
p - b - m	140-480	3	56
t - d - n	130-400	3	75+
k - g	140-400	3	75+
f - v	140-400	3	75
s - z	130-400	3	75+
ʃ - ʒ	150-440	3	75+
θ - ð	140-400	3	75+
l	150-480	3	75+

TABLE 4

List 4 (V̆C/u/JD)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB) High F1	ATTENUATION (dB) Low F2
p - b - m	140-400	3	75
t - d - n	150-480	3	67
k - g	150-400	3	75+
f - v	150-560	3	68
s - z	160-440	3	75+
ʃ - ʒ	160-400	3	75+
θ - ð	150-480	3	75
l	150-400	3	75+

TABLE 5

List 5 (CV/I/RS)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB) High F1	ATTENUATION (dB) Low F2
p - b - m	170-600	3	57
t - d - n	170-600	3	75+
k - g	170-600	3	75+
f - v	170-640	9	50
s - z	140-640	3	75+
ʃ - ʒ	150-600	3	75+
θ - ð	140-680	3	75+
l	140-680	3	75+

TABLE 6

List 6 (VC/I/RS)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB)	
		High F1	Low F2
p - b - m	150-600	3	56
t - d - n	160-600	3	75+
k - g	160-520	3	75+
f - v	160-600	3	75+
s - z	160-640	3	75+
ʃ - ʒ	170-600	3	75+
θ - ð	160-640	3	75+
l	170-600	3	69

TABLE 7

List 7 (CV/u/RS)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB)	
		High F1	Low F2
p - b - m	160-480	3	70
t - d - n	180-480	3	75+
k - g	170-560	3	63
f - v	180-580	3	63
s - z	180-480	3	75+
ʃ - ʒ	180-560	3	75+
θ - ð	160-520	3	75+
l	180-480	3	75+

TABLE 8

List 8 (VC/u/RS)

Consonant Group	Band-Pass Cut-Off Frequencies (Hz)	ATTENUATION (dB) High F1	Low F2
p - b - m	160-520	3	50
t - d - n	180-520	3	70
k - g	180-440	3	65
f - v	170-440	3	75+
s - z	160-480	3	72
ʃ - ʒ	160-480	3	75+
θ - ð	160-480	3	75+
l	190-440	3	75+

APPENDIX D

ANALYSIS OF VARIANCE TABLES

TABLES 1-20

TABLE 1

Summary of the analysis of variance for the number of correct consonant identifications as a function of conditions of filtering (F) and of consonants (C).

Source	df	MS	F
Subjects (S)	19	106.1685	
Conditions of filtering (F)	3	10776.18	570.2113*
Consonants (C)	16	2687.618	77.8629*
S x F	57	18.89836	
S x C	304	34.51731	
F x C	48	280.8401	24.0490*
S x F x C	912	11.67782	

* $p < .01$

TABLE 2

Summary of the analysis of variance for the number of place errors as a function of conditions of filtering (F) and of consonants (C).

Source	df	MS	F
Subjects (S)	19	58.42612	
Conditions of filtering (F)	3	9780.388	679.2465*
Consonants (C)	16	2707.605	84.2604*
S x F	57	14.39888	
S x C	304	32.13376	
F x C	48	352.6959	30.3768*
S x F x C	912	11.61069	

* $p < .01$

TABLE 3

Summary of the analysis of variance for the number of manner errors as a function of conditions of filtering (F) and of consonants (C).

Source	df	MS	F
Subjects (S)	19	76.12674	
Conditions of filtering (F)	3	5118.392	203.6569*
Consonants (C)	16	1415.306	107.3358*
S x F	57	25.13242	
S x C	304	13.18579	
F x C	48	172.5170	26.8069*
S x F x C	912	6.435542	

* $p < .01$

TABLE 4

Summary of the analysis of variance for the number of voicing errors as a function of conditions of filtering (F) and of consonants (C).

Source	df	MS	F
Subjects (S)	19	61.59206	
Conditions of filtering (F)	3	515.2596	54.1848*
Consonants (C)	16	76.10018	10.7485*
S x F	57	9.5093	
S x C	304	7.080059	
F x C	48	41.70435	11.6166*
S x F x C	912	3.590058	

* $p < .01$

TABLE 5

Summary of the analysis of variance for the number of correct consonant identifications in Condition F12L as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	13.82272	
Vowels (V)	1	56.41838	14.6237*
Consonant positions (P)	1	107.8596	28.4249*
Consonants (C)	16	288.6202	73.3552*
S x V	19	3.858011	
S x P	19	3.794544	
V x P	1	88.53603	24.8425*
S x C	304	3.934558	
V x C	16	56.46837	17.2955*
P x C	16	57.03768	18.8638*
S x V x P	19	3.563893	
S x V x C	304	3.264919	
S x P x C	304	3.023655	
V x P x C	16	39.52040	14.2997*
S x V x P x C	304	2.763728	

* $p < .01$

TABLE 6

Summary of the analysis of variance for the number of place errors in Condition F12L as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	7.788351	
Vowels (V)	1	12.23603	4.8480
Consonant positions (P)	1	86.50661	25.9046*
Consonants (C)	16	292.6046	85.6101*
S x V	19	2.523955	
S x P	19	3.339435	
V x P	1	78.62426	26.3539*
S x C	304	3.417875	
V x C	16	49.31260	16.7487*
P x C	16	72.32068	25.5300*
S x V x P	19	2.983398	
S x V x C	304	2.944268	
S x P x C	304	2.832773	
V x P x C	16	34.37583	13.7336*
S x V x P x C	304	2.503052	

* p .01

TABLE 7

Summary of the analysis of variance for the number of manner errors in Condition F12L as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	5.009713	
Vowels (V)	1	63.12426	27.9251*
Consonant positions (P)	1	47.81250	22.2708*
Consonants (C)	16	69.86866	60.5513*
S x V	19	2.260488	
S x P	19	2.14685	
V x P	1	23.03603	16.0484*
S x C	304	1.153875	
V x C	16	23.22582	28.1230*
P x C	16	31.56719	32.0873*
S x V x P	19	1.435410	
S x V x C	304	0.8258659	
S x P x C	304	0.987896	
V x P x C	16	31.07509	36.5816*
S x V x P x C	304	0.8494727	

* $p < .01$

TABLE 8

Summary of the analysis of variance for the number of voicing errors in Condition F12L as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	3.838816	
Vowels (V)	1	21.50074	10.9497*
Consonant positions (P)	1	4.359558	3.7587
Consonants (C)	16	3.626562	6.3008
S x V	19	1.963583	
S x P	19	1.159868	
V x P	1	1.359558	2.1565
S x C	304	0.5755757	
V x C	16	2.446048	5.7286
P x C	16	0.8454963	1.7871
S x V x P	19	0.6304566	
S x V x C	304	0.4269882	
S x P x C	304	0.4731085	
V x P x C	16	0.8142463	2.4177
S x V x P x C	304	0.3367889	

* $p < .01$

TABLE 9

Summary of the analysis of variance for the number of correct consonant identifications in Condition F1 as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	2.909598	
Vowels (V)	1	70.66176	26.5779*
Consonant positions (P)	1	110.6941	65.8214*
Consonants (C)	16	83.55799	18.4733*
S x V	19	2.658669	
S x P	19	1.681733	
V x P	1	21.25000	9.5363*
S x C	304	4.523166	
V x C	16	13.963333	7.5949
P x C	16	37.61756	12.0761*
S x V x P	19	2.228328	
S x V x C	304	1.838520	
S x P x C	304	3.115039	
V x P x C	16	7.032812	4.0849
S x V x P x C	304	1.721667	

* $p < .01$

TABLE 10

Summary of the analysis of variance for the number of place errors in Condition F1 as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	3.135603	
Vowels (V)	1	31.81176	9.4424*
Consonant positions (P)	1	30.59999	21.6845*
Consonants (C)	16	120.3313	22.9462*
S x V	19	3.369040	
S x P	19	1.411145	
V x P	1	29.41176	9.2593*
S x C	304	5.244074	
V x C	16	13.93208	7.3053
P x C	16	45.77655	12.1786*
S x V x P	19	3.176470	
S x V x C	304	1.907115	
S x P x C	304	3.758761	
V x P x C	16	7.607077	3.7847
S x V x P x C	304	2.009940	

* $p < .01$

TABLE 11

Summary of the analysis of variance for the number of manner errors in Condition F1 as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	13.58123	
Vowels (V)	1	71.11838	16.9562*
Consonant positions (P)	1	701.8594	185.1921*
Consonants (C)	16	225.8082	67.1963*
S x V	19	4.194233	
S x P	19	3.789899	
V x P	1	68.40073	16.6877*
S x C	304	3.360424	
V x C	16	20.43244	11.5557*
P x C	16	58.44861	20.6560*
S x V x P	19	4.098877	
S x V x C	304	1.768165	
S x P X C	304	2.829620	
V x P x C	16	12.03355	6.3243
S x V x P x C	304	1.902743	

* $p < .01$

TABLE 12

Summary of the analysis of variance for the number of voicing errors in Condition F1 as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	7.256966	
Vowels (V)	1	10.94412	7.6855*
Consonant positions (P)	1	15.24706	7.0475*
Consonants (C)	16	16.69365	11.6706*
S x V	19	1.423993	
S x P	19	2.163468	
V x P	1	6.497058	5.2131*
S x C	304	1.430403	
V x C	16	10.25193	14.3642*
P x C	16	4.945496	5.1221
S x V x P	19	1.246285	
S x V x C	304	0.7137142	
S x P x C	304	0.9655234	
V x P x C	16	2.542371	3.4026
S x V x P x C	304	0.7471894	

* $p < .01$

TABLE 13

Summary of the analysis of variance for the number of correct consonant identifications in Condition F2 as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	14.03003	
Vowels (V)	1	0.2382353	0.0578
Consonant positions (P)	1	212.8265	36.3057*
Consonants (C)	16	206.6514	44.0314*
S x V	19	4.122136	
S x P	19	5.862074	
V x P	1	17.89411	4.4794
S x C	304	4.693271	
V x C	16	53.78354	16.3599*
P x C	16	43.26241	9.6555*
S x V x P	19	3.994737	
S x V x C	304	3.287514	
S x P x C	304	4.480578	
V x P x C	16	22.47693	8.3627*
S x V x P x C	304	2.687746	

* $p < .01$

TABLE 14

Summary of the analysis of variance for the number of place errors in Condition F2 as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	7.814396	
Vowels (V)	1	1.176470	0.2283
Consonant positions (P)	1	338.0030	81.4465*
Consonants (C)	16	246.4951	59.8805*
S x V	19	5.153251	
S x P	19	4.1500	
V x P	1	43.77646	13.6445*
S x C	304	4.116452	
V x C	16	51.13115	17.5873*
P x C	16	69.27637	17.5394*
S x V x P	19	3.208359	
S x V x C	304	2.907280	
S x P x C	304	3.949753	
V x P x C	16	12.39366	5.4686
S x V x P x C	304	2.266336	

* $p < .01$

TABLE 15

Summary of the analysis of variance for the number of manner errors in Condition F2 as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	12.40290	
Vowels (V)	1	0.4595588	0.3405
Consonant positions (P)	1	72.03603	17.7461*
Consonants (C)	16	63.62114	37.4074*
S x V	19	1.349652	
S x P	19	4.059249	
V x P	1	7.206618	4.3900
S x C	304	1.700764	
V x C	16	12.64081	9.9841*
P x C	16	21.59227	12.3475*
S x V x P	19	1.641602	
S x V x C	304	1.266099	
S x P x C	304	1.748722	
V x P x C	16	6.409743	5.8816
S x V x P x C	304	1.089792	

* $p < .01$

TABLE 16

Summary of the analysis of variance for the number of voicing errors in Condition F2 as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	9.103095	
Vowels (V)	1	2.473529	1.9185
Consonant positions (P)	1	23.29706	5.4669
Consonants (C)	16	26.14862	14.2608*
S x V	19	1.289318	
S x P	19	4.261455	
V x P	1	0.9529411	0.5332
S x C	304	1.833606	
V x C	16	4.684467	5.6270
P x C	16	6.642371	4.3570
S x V x P	19	1.787306	
S x V x C	304	0.8324932	
S x P x C	304	1.524531	
V x P x C	16	12.78263	12.8009*
S x V x P x C	304	0.9985730	

* $p < .01$

TABLE 17

Summary of the analysis of variance for the number of correct consonant identifications in Condition F12H as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	9.953715	
Vowels (V)	1	16.10588	2.6820
Consonant positions (P)	1	431.4838	109.3792*
Consonants (C)	16	303.7049	71.6845*
S x V	19	6.005263	
S x P	19	3.944427	
V x P	1	15.24706	4.2554
S x C	304	4.236692	
V x C	16	83.60744	25.9202*
P x C	16	49.82416	13.5245*
S x V x P	19	3.582972	
S x V x C	304	3.225576	
S x P x C	304	3.683983	
V x P x C	16	35.99237	12.0199*
S x V x P x C	304	2.994403	

* $p < .01$

TABLE 18

Summary of the analysis of variance for the number of place errors in Condition F12H as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	6.667338	
Vowels (V)	1	0.07352938	0.0132
Consonant positions (P)	1	312.5765	97.9075*
Consonants (C)	16	281.9922	71.1552*
S x V	19	5.581270	
S x P	19	3.192569	
V x P	1	38.22353	19.4367*
S x C	304	3.963061	
V x C	16	78.25478	24.7599*
P x C	16	61.22959	17.9481*
S x V x P	19	1.966563	
S x V x C	304	3.160546	
S x P x C	304	3.411484	
V x P x C	16	35.72040	11.4681*
S x V x P x C	304	3.114754	

* $p < .01$

TABLE 19

Summary of the analysis of variance for the number of manner errors in Condition F12H as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	6.887151	
Vowels (V)	1	95.29412	41.3710*
Consonant positions (P)	1	0.04705882	0.0134
Consonants (C)	16	123.9164	64.9443*
S x V	19	2.303405	
S x P	19	3.508359	
V x P	1	1.297058	0.5351
S x C	304	1.908039	
V x C	16	25.95661	17.8269*
P x C	16	38.18769	24.0208*
S x V x P	19	2.423993	
S x V x C	304	1.456037	
S x P x C	304	1.589773	
V x P x C	16	28.29081	21.3413*
S x V x P x C	304	1.325638	

* $p < .01$

TABLE 20

Summary of the analysis of variance for the number of voicing errors in Condition F12H as a function of vowels (V), consonant positions (P) and consonants (C).

Source	df	MS	F
Subjects (S)	19	2.331114	
Vowels (V)	1	12.81176	13.5590*
Consonant positions (P)	1	0.04705882	0.0306
Consonants (C)	16	3.834466	6.1551
S x V	19	0.9448916	
S x P	19	1.536223	
V x P	1	1.838235	2.6776
S x C	304	0.6229731	
V x C	16	3.122702	7.9926
P x C	16	2.148621	4.3871
S x V x P	19	0.6865325	
S x V x C	304	0.3906976	
S x P x C	304	0.4897591	
V x P x C	16	1.902297	4.6947
S x V x P x C	304	0.4052003	

* $p < .01$

APPENDIX E

RANK ORDERED MEAN SCORES AND RESULTS OF TUKEY PROCEDURE FOR HONESTLY SIGNIFICANT DIFFERENCES BETWEEN FILTER CONDITIONS AND BETWEEN CONSONANTS

(ANALYSES 1-20)

TABLES 1-15

TABLE 1

Rank ordered mean scores and significant differences (.01 level) among the four experimental conditions with respect to correct consonant identification, place, manner and voicing errors, as determined by the Tukey procedure.

<u>Correct Consonant Identification</u> (Tukey Range: 1.0856)

Conditions	F1	F2	F12H	F12L
Mean Correct	7.42353	17.90588	18.04118	19.73235
F1	-	10.48235	10.61765	12.30882
F2		-	NS	1.82647
F12H			-	1.69117

<u>Place Errors</u> (Tukey Range: 0.9384)

Conditions	F12L	F2	F12H	F1
Mean Error	10.84412	11.19412	12.42941	22.12941
F12L	-	NS	1.58529	11.28529
F2		-	1.23529	10.93529
F12H			-	9.70000

<u>Manner Errors</u> (Tukey Range: 1.2436)

Conditions	F12L	F2	F12H	F1
Mean Error	3.06765	3.46176	4.25294	11.29118
F12L	-	NS	NS	8.22353
F2		-	NS	7.82942
F12H			-	7.03824

<u>Voicing Errors</u> (Tukey Range: 0.7774)

Conditions	F12H	F12L	F1	F2
Mean Error	1.22941	1.25000	3.11765	3.57647
F12H	-	NS	1.88824	2.34706
F12L		-	4.36765	2.32647
F1			-	NS

TABLE 2

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of correct consonant identifications in Condition F12L. (Tukey Range: 1.2088)

	s	ð	θ	z	v	t	p	l	f	ʒ	m	ʃ	d	b	n	g	k
	1.8375	2.4125	2.5250	3.0125	3.2500	3.3875	4.4625	4.5375	5.0750	5.9625	6.0125	6.1750	6.6875	6.8000	6.9250	7.0750	7.7250
s					1.4125	1.5500	2.6250	2.7000	3.2375	4.1250	4.1750	4.3375	4.8500	4.9600	5.0875	5.2375	5.8875
ð							2.0500	2.1250	2.6625	3.5500	3.6000	3.7625	4.2750	4.3875	4.5125	4.6625	5.3125
θ							1.9375	2.0125	2.5500	3.4375	3.4875	3.6500	4.1625	4.2750	4.4000	4.5500	5.2000
z							1.4500	1.5250	2.0625	2.9500	3.0000	3.1625	3.6750	3.7875	3.9125	4.0600	4.7125
v								1.2875	1.8250	2.7125	2.7625	2.9250	3.4375	3.5500	3.6750	4.4575	4.4750
t									1.6875	2.5750	2.6250	2.7875	3.3000	3.4125	3.5375	3.6875	4.3375
p										1.5000	1.5500	1.7125	2.2250	2.3375	2.4625	2.6125	3.2625
l										1.4250	1.4750	1.6375	2.1500	2.2625	2.3875	2.5375	3.1875
f													1.6125	1.7250	1.8500	2.0000	2.6500
ʒ																	1.7625
m																	1.7125
ʃ																	1.5500
d																	
b																	
n																	
g																	
k																	

TABLE 3

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of place errors in Condition F12L. (Tukey Range: 1.2600)

	k	n	g	m	b	d	ʃ	ʒ	l	f	p	z	t	v	ð	θ	s
	0.2500	0.5250	0.7875	0.8625	1.0500	1.3000	1.4870	1.9875	2.0000	2.5875	3.3875	4.2625	4.4625	4.6875	5.2250	5.2375	5.9875
k							1.2370	1.7375	1.7500	2.3375	3.1375	4.0100	4.2125	4.4375	4.9750	4.9875	5.7375
n								1.4225	1.4750	2.0625	2.8625	3.7400	3.9375	4.1600	4.7000	4.7125	5.4625
g								1.2000	1.2125	1.8000	2.6000	3.4750	3.6750	3.9000	4.4375	4.4500	5.2000
m										1.7250	2.5250	3.4000	3.6000	3.8250	4.3600	4.3750	5.1250
b										1.5375	2.3375	3.2100	3.4100	3.6370	4.1750	4.1875	4.9370
d										1.2875	2.0870	2.9600	3.1600	3.3870	3.9250	3.9400	4.6875
ʃ											1.9000	2.7700	2.9700	3.2000	3.7380	3.7500	4.5005
ʒ											1.4000	2.2700	2.4750	2.7000	3.2375	3.2500	4.0000
l											1.3870	2.2600	2.4625	2.6875	3.2250	3.2375	3.9875
f												1.6750	1.8750	2.1000	2.6375	2.6500	3.4000
p														1.3000	1.8375	1.8500	2.6000
z																	1.7250
t																	1.5250
v																	1.3000
ð																	
θ																	
s																	

TABLE 4

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of manner errors in Condition F12L. (Tukey Range: 0.6540)

	k	d-ʃ	g	p	s	t	b	f-θ	n	z	ʒ	m	ð	v	l
	0.0125	0.6250	0.0750	0.1375	0.1625	0.2125	0.3000	0.5000	0.5875	0.8250	0.9125	1.3250	1.7250	2.1375	3.4625
k										0.8125	0.9000	1.3125	1.7125	2.1250	3.4500
d-ʃ										0.7625	0.8500	1.2625	1.6625	2.0750	3.4000
g										0.7500	0.8375	1.2500	1.6500	2.0625	3.3875
p										0.6875	0.7750	1.1875	1.5875	2.0000	3.3250
s											0.7500	1.1625	1.5625	1.9750	3.3000
t											0.7000	1.1125	1.5125	1.9250	3.2500
b												1.0250	1.4250	1.8375	3.1625
f-θ												0.8250	1.2250	1.6375	2.9625
n												0.7375	1.1375	1.5500	2.8750
z													0.9000	1.3125	2.6375
ʒ													0.8125	1.2250	2.5500
m														0.8125	2.1375
ð															1.7375
v															1.3250
l															

TABLE 5

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of correct consonant identifications in Condition F1. (Tukey Range: 1.2955)

	ʃ	ʒ	θ	ð-z	s	m-l	v	f	g	p	t	k	n	b	d
	0.3750	0.4500	0.6125	0.7125	1.3500	1.8000	1.9250	1.9875	2.1625	2.2625	2.6625	2.8875	3.2250	3.2625	3.3625
ʃ						1.4250	1.5500	1.6125	1.7875	1.8875	2.2875	2.5125	2.8500	2.8875	2.9875
ʒ						1.3500	1.4750	1.5375	1.7125	1.8125	2.2125	2.4375	2.7750	2.8125	2.9125
θ								1.3750	1.5500	1.6500	2.0500	2.2750	2.6125	2.6500	2.7500
ð-z									1.4500	1.5500	1.9500	2.1750	2.5125	2.5500	2.6500
s												1.5375	1.8750	1.9125	2.0125
m-l													1.4250	1.4650	1.5625
v														1.3375	1.4375
f															1.3750
g															
p															
t															
k															
n															
b															
d															

TABLE 6

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of place errors in Condition F1. (Tukey Range: 1.3947)

	n	l	b	d	t	z	s	k	m	p	f	g	v	ð	θ	ʒ	ʃ
	3.7500	3.8125	4.4875	4.6000	4.6500	4.7625	4.9625	5.0250	5.5250	5.5375	5.6500	5.7625	6.0500	7.0125	7.3625	7.5375	7.5625
n									1.7750	1.7875	1.9000	2.0125	2.3000	3.2615	3.6125	3.7875	3.8125
l									1.7125	1.7250	1.8375	1.9500	2.2375	3.2000	3.5500	3.7250	3.7500
b													1.5625	2.5250	2.8750	3.0500	3.0750
d													1.4500	2.4125	2.7625	2.9375	2.9625
t														2.3625	2.7125	2.8875	2.9125
z														2.2500	2.6000	2.7750	2.8000
s														2.0500	2.4000	2.5750	2.6000
k														1.9875	2.3375	2.5125	2.5375
m														1.4875	1.8375	2.0125	2.0375
p														1.4750	1.8250	2.0000	2.0250
f															1.7125	1.8875	1.9125
g															1.6000	1.7750	1.8000
v																1.4875	1.5125
ð																	
θ																	
ʒ																	
ʃ																	

TABLE 7

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of manner errors in Condition F1. (Tukey Range: 1.1167)

	d	g	k	b	p	t	n	ʃ	s	m	θ	f	v	ð	z	ʒ	l
	0.1750	0.5000	0.7500	1.1375	1.3250	1.8125	2.3125	2.5875	3.1370	3.3750	3.8250	4.0000	4.0375	4.0500	4.3750	4.3875	6.2000
d					1.1500	1.5375	2.1375	2.4125	2.9620	3.2000	3.6500	3.8250	3.8625	3.8750	4.2000	4.2125	6.0250
g						1.2125	1.8125	2.0875	2.6370	2.8750	3.3250	3.5000	3.5375	3.5500	3.8750	3.8875	5.7000
k							1.5625	1.8375	2.3870	2.6250	3.0750	3.2500	3.2875	3.3000	3.6250	3.6375	5.4500
b							1.1750	1.4500	1.9995	2.2375	2.6875	2.8625	2.9000	2.9125	3.2375	3.2500	5.0625
p								1.2625	1.8120	2.0500	2.5000	2.6750	2.7125	2.7250	3.0500	3.0625	4.8750
t									1.3245	1.5625	2.0125	2.1875	2.2250	2.2375	2.5625	2.5750	4.3875
n											1.5125	1.6875	1.7350	1.7475	2.0725	2.0750	3.8875
ʃ											1.2375	1.4125	1.4500	1.4625	1.7875	1.8000	3.6125
s															1.2380	1.2505	3.0630
m																	2.8250
θ																	2.3750
f																	2.2000
v																	2.1625
ð																	2.1500
z																	1.8250
ʒ																	1.8125
l																	

TABLE 8

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of voice errors in Condition F1. (Tukey Range: 0.7281)

	d	g	n	z	v	m	ð	k	l	b	ʒ	t	θ	p	f	ʃ	s
	0.1250	0.2000	0.3375	0.3500	0.5500	0.5750	0.5875	0.6250	0.7000	0.7375	0.7500	0.9750	1.0125	1.1125	1.4000	1.4125	1.8000
d												0.8500	0.8875	0.9875	1.2750	1.2875	1.6750
g												0.7750	0.8125	0.9125	1.2000	1.2125	1.6000
n														0.7750	1.0625	1.0750	1.4625
z														0.7625	1.0500	1.0625	1.4500
v															0.8500	0.8625	1.2500
m															0.8250	0.8375	1.2250
ð															0.8125	0.8250	1.2125
k															0.7750	0.7875	1.1750
l																	1.1000
b																	1.0625
ʒ																	1.0500
t																	0.8250
θ																	0.7875
p																	
f																	
ʃ																	
s																	

TABLE 9

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of correct consonant identifications in Condition F2. (Tukey Range: 1.3193)

	ð	θ	s	z	p	t	g	l	v	b	f	d	ʒ	n	ʃ	m	k
	1.9750	2.0500	2.4000	3.2875	3.4750	3.8250	4.1375	4.2125	4.2625	4.3750	4.5625	5.3125	5.7875	6.0375	6.2750	6.6625	7.4625
ð					1.5000	1.8500	2.1625	2.2375	2.2875	2.4000	2.5875	3.3375	3.8125	4.0625	4.3000	4.6875	5.4875
θ					1.4250	1.7750	2.0875	2.1625	2.2125	2.3250	2.5125	3.2625	3.7375	3.9875	4.2250	4.6125	5.4125
s						1.4250	1.7375	1.8125	1.8625	1.9750	2.1625	2.9125	3.3875	3.6375	3.8750	4.2625	5.0625
z												2.0250	2.5000	2.7500	2.9875	3.3750	4.1750
p												1.8375	2.3125	2.5625	2.8000	3.1875	3.9875
t												1.4875	1.9625	2.2125	2.4500	2.8375	3.6375
g													1.6500	1.9000	2.1375	2.5250	3.3250
l													1.5750	1.8250	2.0625	2.4500	3.2500
v													1.5250	1.7750	2.0125	2.4000	3.2000
b													1.4125	1.6625	1.9000	2.2875	3.0875
f														1.4750	1.7125	2.1000	2.9000
d																	2.1500
ʒ																	1.6750
n																	1.4250
ʃ																	
m																	
k																	

TABLE 10

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of place errors in Condition F2. (Tukey Range: 1.2355)

	k	ʃ	n	ʒ	m	d	g	b	l	f	v	t	p	z	s	ð	θ
	0.3875	0.9250	0.9500	0.9750	1.0125	1.6000	2.0375	2.1750	2.5625	2.8250	3.2500	3.9375	4.1750	4.2500	5.2250	5.6125	5.6750
k							1.6500	1.7875	2.1750	2.4375	2.8625	3.5500	3.7875	3.8600	4.8375	5.2250	5.2875
ʃ									1.6375	1.9000	2.3250	3.0125	3.2500	3.3250	4.3000	4.6875	4.7500
n									1.6125	1.8750	2.3000	2.9875	3.2250	3.3000	4.2750	4.6625	4.7250
ʒ									1.5875	1.8500	2.2750	2.9625	3.2000	3.2750	4.2500	4.6375	4.7000
m									1.5500	1.8125	2.2375	2.9250	3.1600	3.2375	4.2125	4.6000	4.6625
d											1.6500	2.3375	2.5750	2.6500	3.6250	4.0125	4.0750
g												1.9000	2.1375	2.2125	3.1875	3.5750	3.6375
b												1.7625	2.0000	2.0750	3.0500	3.4375	3.5000
l												1.3750	1.6125	1.6875	2.6625	3.0500	3.1125
f													1.3500	1.4250	2.4000	2.7875	2.8500
v															1.9750	2.3625	2.4250
t															1.2875	1.6750	1.7375
p																1.4375	1.5000
z																1.3625	1.4250
s																	
ð																	
θ																	

TABLE 11

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of manner errors in Condition F2. (Tukey Range: 0.7935)

	k	ʃ	z	p	s	ʒ	t	g	θ	v	m	f	d	ð	n	b	l
	0.0625	0.1000	0.1750	0.3250	0.3500	0.4000	0.4875	0.6125	0.6625	0.7000	0.7375	0.8250	1.1000	1.1625	1.3375	1.8875	3.7875
k													1.0375	1.1000	1.2750	1.8250	3.7250
ʃ													1.0000	1.0625	1.2375	1.7875	3.6875
z													0.9250	0.9875	1.1625	1.7125	3.6125
p														0.8375	1.0125	1.5625	3.4600
s															0.9875	1.5375	3.4375
ʒ															0.9375	1.4875	3.3875
t															0.8500	1.4000	3.3000
g																1.2750	3.1750
θ																1.2250	3.1250
v																1.1875	3.0875
m																1.1500	3,0500
f																1.0625	2.9625
d																	2.6875
ð																	2.6250
n																	2.4500
b																	1.0000
l																	

TABLE 12

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of voice errors in Condition F2. (Tukey Range: 0.8246)

	m	n	k	l	s	ð	t	p	d	z	ʃ	θ	v	ʒ	f	b	g
	0.0250	0.1625	0.1875	0.4750	0.5875	0.6000	0.7000	0.7250	0.9375	0.9625	0.9750	1.0500	1.1000	1.4375	1.4750	1.5250	2.2275
m									0.9125	0.9375	0.9500	1.0250	1.0750	1.4125	1.4500	1.5000	2.2025
n												0.8875	0.9375	1.2750	1.3125	1.3625	2.0650
k												0.8625	0.9125	1.2500	1.2875	1.3375	2.0400
l														0.9625	1.0000	1.0500	1.7525
s														0.8500	0.8875	0.9375	1.6400
ð															0.8750	0.9250	1.6275
t																	1.5275
p																	1.5025
d																	1.2900
z																	1.2650
ʃ																	1.2525
θ																	1.1775
v																	1.1275
ʒ																	
f																	
b																	
g																	

TABLE 13

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of correct consonant identifications in Condition F12H. (Tukey Range: 1.2535)

	s	z	ð	θ	v	t	l	f	p	m	ʒ	n	d-ʃ	g	b	k
	1.0875	1.8000	2.1750	2.5000	2.7250	3.3875	3.7750	4.6875	4.7000	5.0500	5.5625	6.2125	6.2750	6.3750	6.6625	7.4250
s				1.4125	1.6375	2.3000	2.6875	3.6000	3.6125	3.9625	4.4750	5.1250	5.1875	5.2875	5.5750	6.3375
z						1.5875	1.9750	2.8875	2.9000	3.2500	3.7625	4.4125	4.4750	4.5750	4.8625	5.6250
ð							1.6000	2.5125	2.5250	2.8750	3.3875	4.0375	4.1000	4.2000	4.4875	5.2500
θ								2.1875	2.2000	2.5500	3.0625	3.7125	3.7750	3.8750	4.1625	4.9250
v								1.9625	1.9750	2.3250	2.8375	3.4875	3.5500	3.6500	3.9375	4.7000
t								1.3000	1.3125	1.6625	2.1750	2.8250	2.8875	2.9875	3.2750	4.0375
l											1.7875	2.4375	2.5000	2.6000	2.8875	3.6500
f												1.5250	1.5875	1.6875	1.9750	2.7375
p												1.5125	1.5750	1.6750	1.9625	2.7250
m														1.3250	1.6125	2.3750
ʒ																1.8625
n																
d-ʃ																
g																
b																
k																

TABLE 14

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the number of place errors in Condition F12H. (Tukey Range: 1.2126)

	k	b	n	g	m	d-ʃ	ʒ	l	f	p	t	z	v	θ	ð	s
	0.5625	1.2125	1.2625	1.4250	1.5375	1.6625	2.3875	2.8375	3.0250	3.2125	4.5875	4.7250	5.2125	5.3750	5.3870	6.7500
k							1.8250	2.2750	2.4625	2.6500	4.0250	4.1625	4.6500	4.8125	4.8245	6.1875
b								1.6250	1.8125	2.0000	3.3750	3.5125	4.0000	4.1625	4.1745	5.5375
n								1.5750	1.7625	1.9500	3.3250	3.4625	3.9500	4.1125	4.1245	5.4875
g								1.4125	1.6000	1.7875	3.1625	3.3000	3.7875	3.9500	3.9620	5.3250
m									1.4875	1.6750	3.0500	3.1875	3.6750	3.8375	3.8495	5.2125
d-ʃ									1.3625	1.5500	2.9250	3.0625	3.5500	3.7125	3.7245	5.0875
ʒ											2.2000	2.3375	2.8250	2.9875	2.9995	4.3625
l											1.7500	1.8875	2.3750	2.5375	2.5495	3.9125
f											1.5625	1.7000	2.1875	2.3500	2.3620	3.7250
p											1.3750	1.5125	2.0000	2.1625	2.1745	3.5375
t																2.1625
z																2.0250
v																1.5375
θ																1.3750
ð																1.3630
s																

TABLE 15

Rank ordered consonant mean scores and significant differences among the 17 consonants with respect to the **number of manner** errors in Condition F12H. (Tukey Range: 08409)

	d-k	ʃ	p-t	s	g	b	θ	n	f	ʒ	z	m	ð	v	l
	0.0000	0.0375	0.1125	0.2000	0.2875	0.3000	0.4000	0.7625	1.0500	1.4875	1.7500	1.8875	2.3375	3.1250	4.2250
d-k									1.0500	1.4875	1.7500	1.8875	2.3375	3.1250	4.2250
ʃ									1.0125	1.4500	1.7125	1.8500	2.3000	3.0875	4.1875
p-t									0.9375	1.3750	1.6375	1.7750	2.2250	3.0125	4.1125
s										1.2875	1.5500	1.6875	2.1375	2.9250	4.0250
g										1.2000	1.4625	1.6000	2.0500	2.8375	3.9375
b										1.1875	1.4500	1.5875	2.0375	2.8250	3.9250
θ										1.0875	1.3500	1.4875	1.9375	2.7250	3.8250
n											0.9875	1.1250	1.5750	2.3625	3.4625
f													1.2875	2.0750	3.1750
ʒ														1.6375	2.7375
z														1.3750	2.4750
m														1.2375	2.3375
ð															1.8875
v															
l															

APPENDIX F

DIFFERENTIAL FILTER EFFECTS

PERCENTAGE DIFFERENCES BETWEEN ALL CONDITIONS

FOR EACH CONSONANT

WITH RESPECT TO THE FOUR DEPENDENT VARIABLES

TABLES 1-4

TABLE 1

Percentage differences between conditions for each consonant with respect to correct identifications.

(Tukey Range: 10.97%)

	F12L-F12H	F12L-F2	F12L-F1	F2-F1	F12H-F2	F12H-F1
p	- 3.0	12.4*	27.5*	15.1*	15.4*	30.5*
b	1.7	28.6*	44.2*	13.9*	28.6*	42.5*
t	0.0	- 5.5	9.0	14.5*	- 5.5	9.0
d	5.2	19.2*	41.6*	24.4*	12.0*	36.4*
k	3.8	3.3	60.5*	57.2*	- 0.5	56.7*
g	8.7	36.7*	61.4*	24.7*	28.0*	52.7*
f	4.8	6.4	38.6*	32.2*	1.6	33.8*
v	6.5	-12.7*	16.5*	29.2*	-19.2*	10.0
s	9.4	- 7.0	6.1	13.1*	-16.4*	- 3.3
z	15.2*	- 3.4	28.8*	32.2*	-18.6*	13.6*
ʃ	1.2	- 1.2	72.5*	73.7*	0.0	73.7*
ʒ	5.0	2.2	68.9*	66.7*	- 2.8	63.9*
θ	0.3	6.0	23.9*	17.9*	5.7	23.6*
ð	3.0	5.5	21.3*	15.8*	2.5	18.3*
m	12.1*	- 8.1	52.7*	60.8*	-20.2*	40.6*
n	8.9	11.1*	46.3*	35.2*	2.2	37.4*
l	9.5	4.0	34.2*	30.2*	- 5.5	24.7*

* $p < .01$

TABLE 2

Percentage differences for each consonant with respect to place errors. (Tukey Range: 10.81%)

	F12H-F12L	F2-F12L	F1-F12L	F1-F2	F2-F12H	F1-F12H
p	- 2.1	9.9	26.9*	17.0*	12.0*	29.0*
b	2.1	14.1*	43.0*	28.9*	12.0*	40.9*
t	1.5	- 6.6	2.3	8.9	- 8.1	0.8
d	4.5	3.7	41.2*	37.5*	- 0.8	36.7*
k	3.9	1.7	59.7*	58.0*	- 2.2	55.8*
g	8.0	15.7*	62.2*	46.5*	7.7	54.2*
f	5.5	3.0	38.3*	35.3*	- 2.5	32.8*
v	6.6	-18.0*	17.0*	30.6*	-24.6*	10.4
s	9.6	- 9.5	-12.8*	- 3.3	-19.1*	-22.4*
z	5.8	- 0.2	6.2	6.4	- 6.0	0.4
ʃ	2.2	- 7.0	75.9*	82.9*	- 9.2	73.7*
ʒ	5.0	-12.6*	69.4*	82.0*	-17.8*	64.4*
θ	1.7	5.4	26.5*	21.1*	3.7	24.8*
ð	2.0	4.9	22.4*	17.5*	2.9	20.4*
m	8.4	1.9	58.3*	56.4*	- 6.5	49.9*
n	9.2	5.3	40.3*	35.0*	- 3.9	31.1*
l	10.5	7.0	22.7*	15.7*	35.5*	12.2*

* $p < .01$

TABLE 3

Percentage differences for each consonant with respect to manner errors. (Tukey Range: 8.84%)

	F12H-F12L	F2-F12L	F1-F12L	F1-F2	F2-F12H	F1-F12H
p	- 0.3	2.4	14.9*	12.5*	2.7	15.2*
b	0.0	19.8*	10.4*	- 9.4*	19.8*	10.4*
t	- 1.3	3.4	20.0*	16.6*	4.7	21.3*
d	- 0.8	13.0*	1.4	-11.6*	13.8*	2.2
k	- 0.2	0.6	9.2*	8.6	0.8	9.4*
g	2.7	6.8	5.4	- 1.4	4.1	2.7
f	6.8	4.0	43.7*	39.7*	- 2.8	36.9*
v	12.4*	-17.9*	23.8*	41.7*	-30.3*	11.4*
s	0.5	2.4	37.2*	34.8*	1.9	36.7*
z	11.6*	- 8.1	44.4*	52.5*	-19.7*	32.8*
ʃ	- 0.3	0.5	31.5*	31.0*	0.8	31.8*
ʒ	7.2	- 6.4	43.4*	49.8*	-13.6*	36.2*
θ	- 1.3	2.0	41.5*	39.5*	3.3	42.8*
ð	7.2	- 7.5	28.6*	36.1*	-14.7*	21.4*
m	7.0	- 7.4	25.6*	33.0*	-14.4*	18.6*
n	2.2	9.4*	21.6*	12.2*	7.2	19.4*
l	9.5*	4.2	34.2*	30.0*	- 5.3	24.7*

* p <.01

TABLE 4

Percentage differences for each consonant with respect to voicing errors. (Tukey Range: 6.28%)

	F12H-F12L	F2-F12L	F1-F12L	F1-F2	F2-F12H	F1-F12H
p	- 2.0	4.3	9.1*	4.8	6.3*	11.1*
b	1.0	17.8*	7.9*	- 9.9*	16.8*	6.9*
t	- 2.2	5.0	8.4*	3.4	7.2*	10.6*
d	0.1	10.9*	0.8	-10.1*	10.8*	0.7
k	- 1.1	1.0	6.5*	5.5	2.1	7.6*
g	0.7	26.4*	0.5	-25.9*	25.7*	- 0.2
f	2.0	10.6*	9.7*	- 0.9	8.6*	7.7*
v	1.0	8.5*	1.6	- 6.9*	7.5*	0.6
s	- 0.1	1.7	16.9*	15.2*	1.8	17.0*
z	- 1.1	5.4	- 2.2	- 7.6*	6.5*	- 1.1
ʃ	- 3.4	6.1	11.6*	5.5	9.5*	15.0*
ʒ	1.4	16.3*	7.7*	- 8.6*	14.9*	6.3*
θ	- 3.1	4.7	4.3	- 0.4	7.8*	7.4*
ð	0.9	1.4	1.2	- 0.2	0.5	0.3
m	0.4	- 0.6	6.3*	6.9*	- 1.0	5.9
n	0.9	1.8	4.0	2.2	0.9	3.1
l	3.5	2.1	5.0	2.9	- 1.4	1.5

* $p < .01$

APPENDIX G

SIGNIFICANT VOWEL-BY-POSITION (V x P) AND VOWEL-BY-POSITION-BY-CONSONANT (V x P x C) INTERACTIONS

The dependent variables for which the vowel-by-position and the vowel-by-position-by-consonant interactions were significant are shown in Table 1.

The vowel-by-position interaction was significant for all dependent variables except voicing errors in Conditions F12L and F1, and for place errors only in Conditions F2 and F12H. In Condition F12L, with respect to intelligibility and to the perception of place and manner, large differences in mean scores were noted between the two vowel environments for final consonants, but only small differences for initial consonants. Consonants associated with the vowel /I/ showed a greater increase in intelligibility and decrease in place errors in the final position than those associated with the vowel /u/, for which there was little difference in mean scores between the two positions. More place errors were noted for consonants with /I/ than for those with /u/ in the initial position. With respect to manner errors, there was a slight increase in errors for final consonants in the /I/ environment and a much larger increase for consonants in the /u/ environment. Condition F1 also showed large differences in mean scores between vowels for final consonants and small differences in mean scores between vowels for initial consonants, with respect to intelligibility, place and manner. Consonants associated with the vowel /u/ showed a greater decrease in intelligibility for the final position together with a greater increase in place and manner errors. In Conditions F2 and F12H, consonants in the /I/ environment gave more place errors than those in the /u/ environment for the initial position, but fewer place errors for the final position.

The vowel-by-position-by-consonant interaction was significant for all dependent variables except voicing errors in Conditions F12L

TABLE 1

Significant V x P and V x P x C interactions.

Dependent Variable	FILTER CONDITIONS							
	F12L		F1		F2		F12H	
	V x P	V x P x C	V x P	V x P x C	V x P	V x P x C	V x P	V x P x C
Consonant Identifications	x	x	x			x		x
Place Errors	x	x	x		x		x	x
Manner Errors	x	x	x					x
Voicing Errors						x		

and F12H, and for consonant identifications and voicing errors in Condition F2. In general, these three-way interactions occurred because the vowel-by-position interactions described in the preceding paragraph, and the vowel-by-consonant and position-by-consonant interactions did not occur to the same extent for all consonants. This is illustrated in Figure 1 for the intelligibility scores of the consonants /p/ and /b/ in the two vowel environments and two consonant positions in Condition F12L (Analysis 13). The consonant /p/ was more intelligible in the /I/ environment and in the final position. These same trends are apparent for each position and for each vowel as seen in Figure 1. The consonant /b/ showed no difference in intelligibility in the two vowel environments, and in the two positions, but did show the trends described for the vowel-by-position interaction. It is beyond the scope of this paper to go into the details of these interactions.

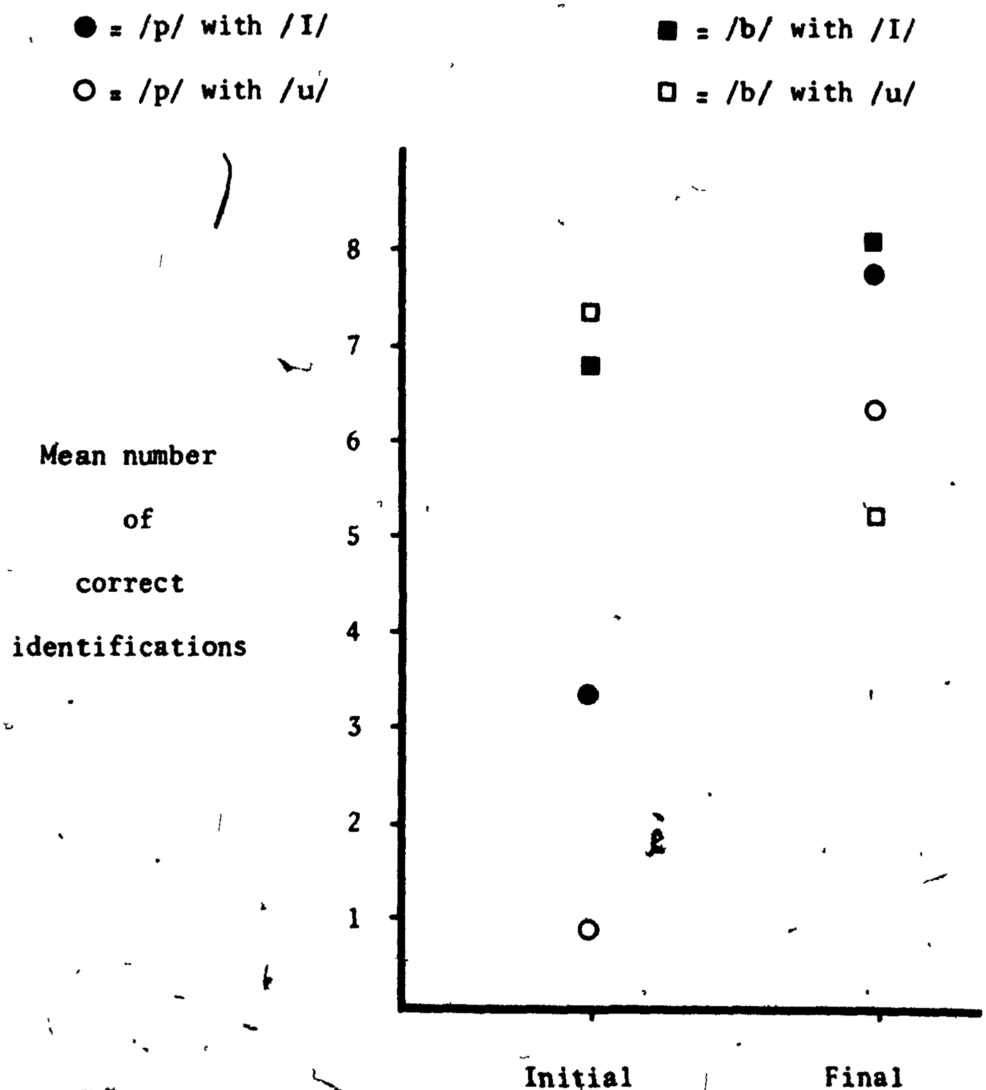

Figure 1: Mean scores for the consonants /p/ and /b/ in the two vowel environments and two positions with respect to correct consonant identifications in Condition F12L. (Maximum score: 8)

APPENDIX H

CONFUSION MATRICES FOR THE 17 CONSONANTS AND FOUR FILTER CONDITIONS

TABLES 1-4

TABLE 1.

Confusion matrix for the consonants in Condition F12L.

	p	b	t	d	k	g	f	v	s	z	ʃ	ʒ	θ	ð	m	n	l
p	357	11	65	2	182	12	0	1	0	0	0	0	5	3	1	1	0
b	4	544	0	25	1	42	1	7	0	0	1	0	1	3	8	2	1
t	165	5	271	5	170	7	4	2	4	0	0	0	2	1	1	3	0
d	1	87	1	535	0	11	2	1	0	0	0	0	1	1	0	0	0
k	2	2	12	3	618	2	0	0	0	1	0	0	0	0	0	0	0
g	1	3	0	53	11	566	0	0	0	1	0	2	1	0	0	1	1
f	13	1	6	0	11	2	406	27	21	4	1	0	132	9	2	3	2
v	1	61	0	6	5	64	5	260	4	47	0	1	19	133	17	10	7
s	3	0	2	1	1	1	208	10	147	7	35	2	208	10	1	4	0
z	1	7	0	20	0	10	0	104	13	241	0	41	28	147	3	16	9
ʃ	2	0	1	2	0	0	23	1	37	7	494	27	44	2	0	0	0
ʒ	0	0	1	10	0	3	3	40	0	33	4	477	3	7	1	43	15
θ	16	2	10	7	1	2	251	12	104	8	2	2	202	19	0	0	2
ð	0	18	0	20	0	11	1	146	9	121	0	0	29	193	3	83	6
m	1	89	0	3	1	2	1	4	3	0	0	0	0	1	481	53	1
n	0	0	0	4	1	2	0	0	0	0	0	0	0	0	39	554	40
l	5	36	2	30	1	6	2	41	5	4	0	2	9	7	51	76	363

TABLE 2

Confusion matrix for the consonants in Condition F1.

	p	b	t	d	k	g	f	v	s	z	ʃ	ʒ	θ	ð	m	n	l
p	181	10	186	16	120	21	47	8	6	4	1	0	10	4	6	17	3
b	12	261	11	221	4	40	18	25	8	4	0	1	6	6	8	7	8
t	139	15	213	8	114	6	48	11	23	3	1	2	24	6	6	16	5
d	5	293	1	269	2	56	1	6	0	0	0	0	1	4	0	1	1
k	108	5	212	17	231	7	26	5	3	2	1	0	9	2	3	8	1
g	3	163	3	252	6	173	0	15	0	3	0	0	4	7	6	4	1
f	89	13	118	10	46	4	159	29	56	4	10	3	50	9	10	22	8
v	6	68	4	63	3	54	2	154	13	71	0	18	16	43	24	70	31
s	59	16	70	9	19	3	186	27	108	5	9	2	45	7	24	38	13
z	3	29	1	82	3	89	8	138	3	57	1	30	9	44	27	63	53
ʃ	55	7	69	11	21	8	208	31	96	7	30	5	48	8	6	19	11
ʒ	6	36	11	63	3	55	6	127	9	55	1	36	24	31	57	73	47
θ	93	8	135	10	26	9	173	19	73	5	10	3	49	2	7	12	6
ð	5	69	3	69	3	30	4	141	7	73	3	9	22	57	36	85	15
m	10	44	4	12	2	7	9	41	10	23	0	10	11	16	144	226	71
n	4	39	7	33	1	12	6	19	4	11	1	4	4	13	197	258	27
l	12	63	18	16	3	22	13	30	6	19	0	9	4	8	141	132	144

TABLE 3

Confusion matrix for the consonants in Condition F2.

	p	b	t	d	k	g	f	v	s	z	ʃ	ʒ	θ	ð	m	n	l
p	278	24	53	1	236	22	10	2	1	0	0	0	4	3	4	0	2
b	56	350	1	9	34	39	22	43	6	0	0	0	3	8	60	5	4
t	82	5	306	9	180	19	9	3	2	0	1	9	4	2	1	7	1
d	8	57	42	425	5	15	4	14	8	5	3	5	5	7	5	24	8
k	3	0	23	0	597	12	1	0	0	0	1	2	0	1	0	0	0
g	11	18	9	76	146	331	3	6	7	2	0	8	6	4	1	4	8
f	8	4	1	1	12	4	365	49	53	8	2	1	81	15	35	1	0
v	0	9	2	10	1	4	39	341	6	36	2	2	38	120	20	3	7
s	2	0	3	0	3	2	220	4	192	18	38	2	135	3	9	9	0
z	1	2	0	3	0	1	8	123	30	263	4	33	34	131	3	4	0
ʃ	0	0	2	1	0	0	7	1	35	5	502	64	16	2	1	4	0
ʒ	1	0	3	2	0	1	4	8	6	24	99	463	2	2	0	20	5
θ	8	4	10	6	3	0	233	7	132	18	6	5	164	22	16	6	0
ð	0	7	0	4	0	1	2	187	13	147	0	7	33	158	26	54	1
m	2	24	0	1	0	0	0	5	0	0	0	1	0	0	533	48	26
n	0	1	5	7	0	1	1	2	1	18	0	0	6	15	50	483	50
l	3	10	1	8	1	3	7	48	6	18	0	2	20	60	51	65	337

TABLE 4

Confusion matrix for the consonants in Condition F12H.

	p	b	t	d	k	g	f	v	s	z	ʃ	ʒ	θ	ð	m	n	l
p	376	7	23	1	217	7	2	0	0	0	0	1	4	0	0	1	1
b	7	533	2	34	2	38	2	13	1	0	0	0	1	1	3	1	2
t	120	2	271	1	233	4	2	0	0	0	1	0	3	2	0	0	1
d	1	109	5	502	0	23	0	0	0	0	0	0	0	0	0	0	0
k	11	0	34	0	594	1	0	0	0	0	0	0	0	0	0	0	0
g	0	27	0	64	16	510	1	12	0	1	0	7	0	2	0	0	0
f	22	5	10	2	29	7	375	23	40	3	12	4	89	10	2	3	4
v	1	79	2	13	4	92	5	218	6	25	1	9	21	105	14	13	32
s	1	0	2	2	6	0	260	3	87	5	68	3	181	17	1	2	2
z	1	11	2	25	0	11	0	177	8	144	0	50	24	157	7	50	33
ʃ	0	0	1	0	1	0	50	1	18	1	502	5	51	9	0	1	0
ʒ	0	0	2	4	0	4	1	32	6	11	4	445	7	15	11	79	19
θ	7	0	15	3	7	0	286	11	79	4	12	6	200	10	0	0	0
ð	0	29	1	30	0	11	4	140	5	84	0	11	35	174	11	86	19
m	3	110	2	1	0	11	1	4	1	1	0	0	1	2	404	85	14
n	0	2	1	9	0	5	0	1	3	4	0	1	3	7	82	497	25
l	12	42	3	11	1	6	14	48	7	1	0	1	10	18	75	89	302

APPENDIX J

THREE TWO-DIMENSIONAL PLOTS
OF THE 17 CONSONANTS IN CONDITION F12L
FROM WHICH THREE-DIMENSIONAL PLOT (FIGURE 6) WAS DERIVED

FIGURES 1-3

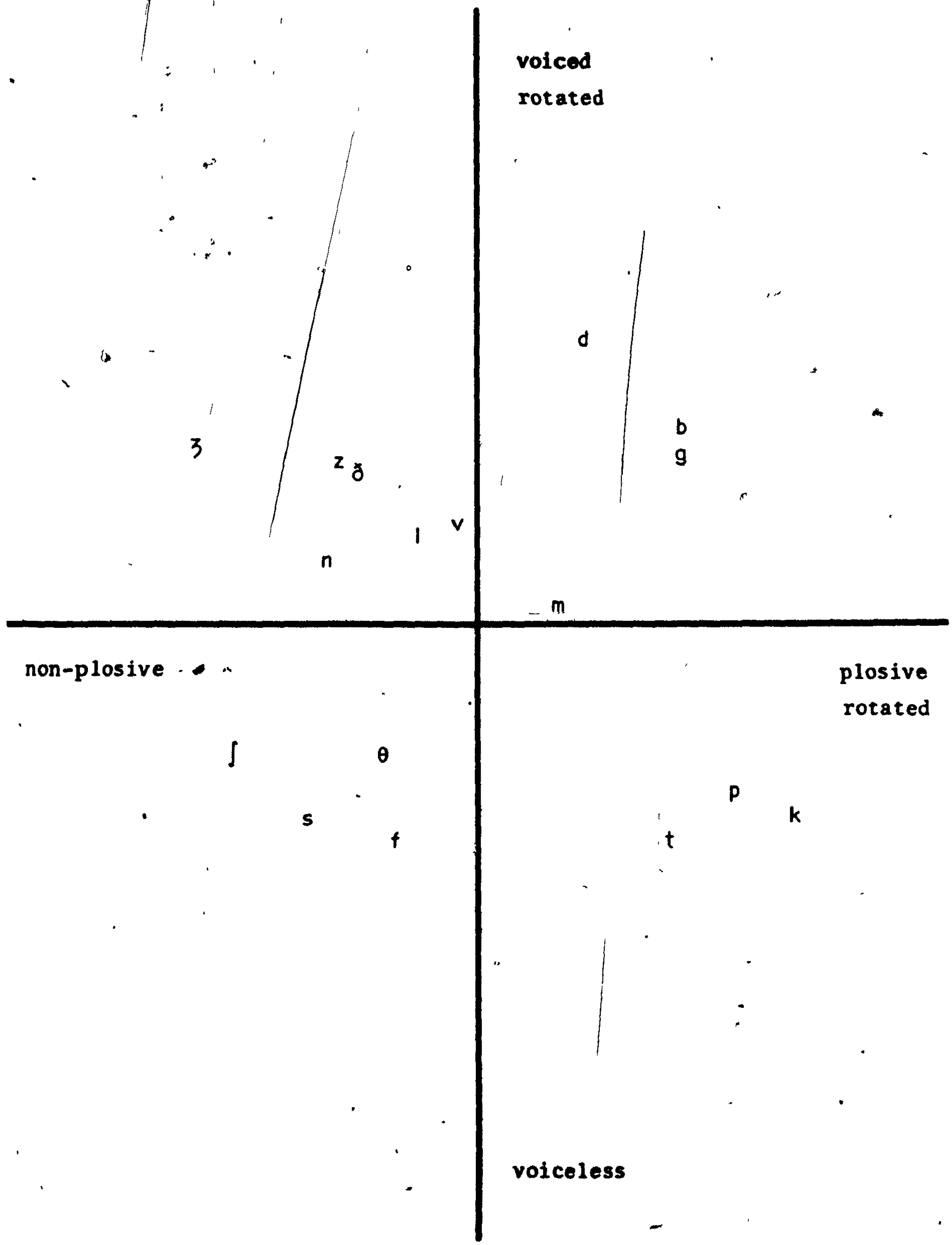

Figure 1: Spatial representation of the 17 consonants for plosive/non-plosive and voiced/voiceless dimensions of Condition F12L.

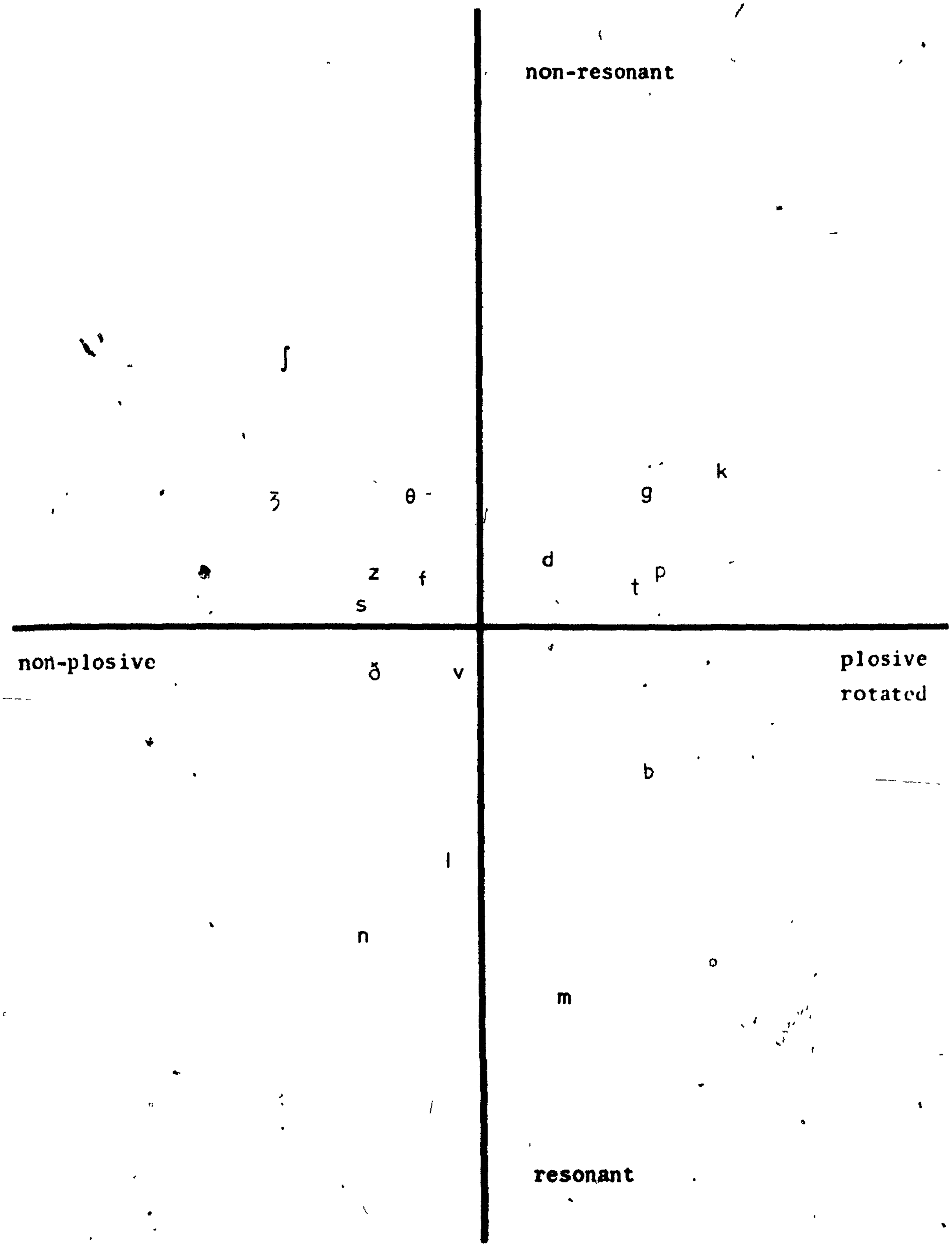

Figure 2: Spatial representation of the 17 consonants for the plosive/non-plosive and resonant/non-resonant dimensions of Condition F12L.

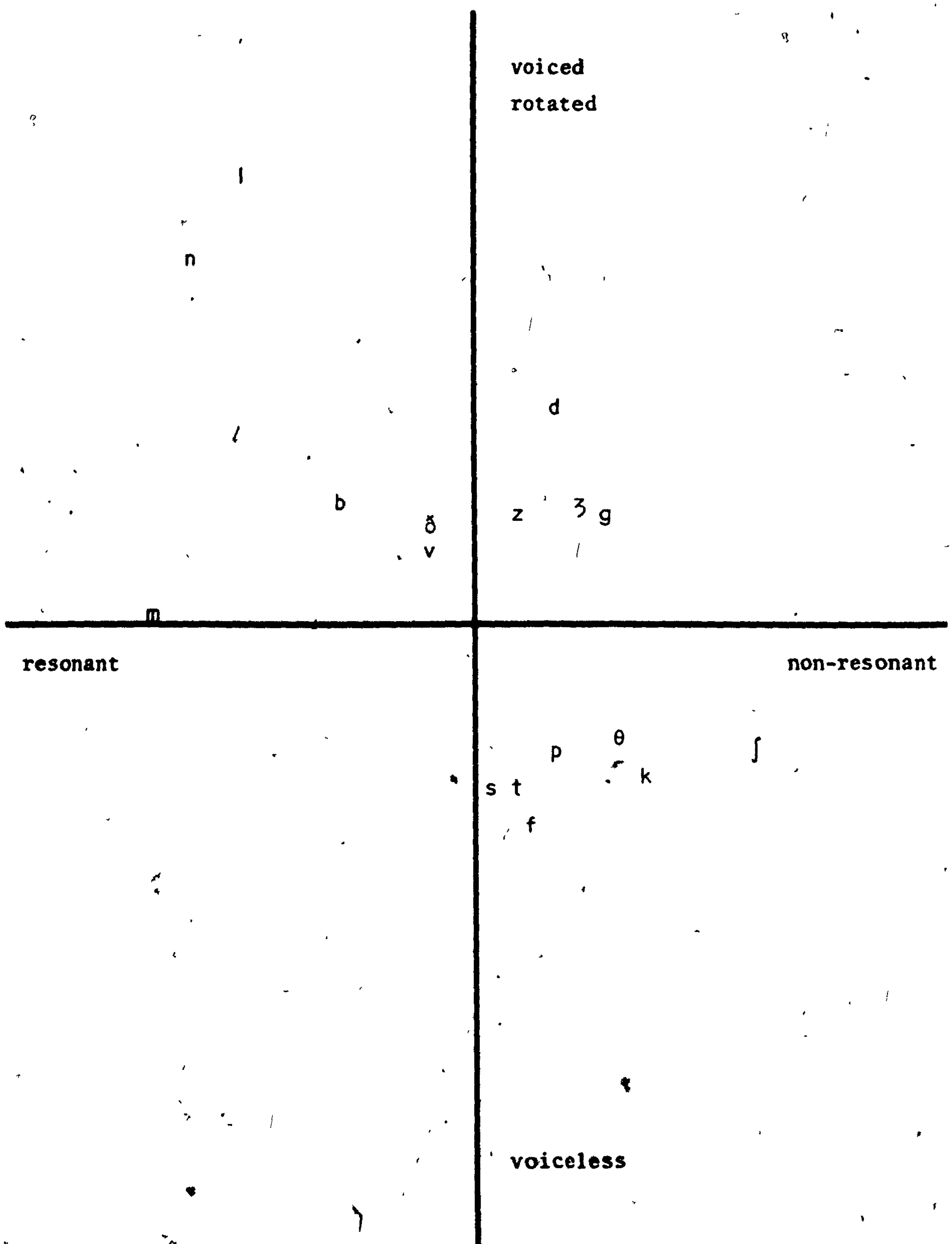

Figure 3: Spatial representation of the 17 consonants for the voiced/voiceless and resonant/non-resonant dimensions of Condition F12L.