

Perception of Features in the Identification of English Consonants

Tobey Lynn Doeleman

Perception of voicing, manner, and place of articulation was investigated in two gating experiments. In Experiment 1, subjects identified the English consonants [p, t, b, d, f, s, v, z] presented in six different gated conditions. Results were analyzed in terms of correct identification of the consonants and each of the features, and in terms of information transmission. In Experiment 2, subjects were presented with two mutually exclusive sets of the same stimuli such that information about a particular feature was given (e.g. voicing given in sets [p, t, f, s] and [b, d, v, z]). The results showed that the subjects given place information had significantly higher identification scores than the subjects given manner information or the subjects in Experiment 1. This suggests a privileged status for place information such that it can reduce ambiguities in voicing and manner and facilitate identification of these consonants.

1 Introduction

Consonants can be described in terms of the linguistic features that distinguish them in a given language. Thus the phoneme [p], for example, can be described in terms of its voicing, its place of articulation, and its manner of articulation as a voiceless, bilabial stop consonant. The phonological representation of each phoneme is thus made up of multiple featural specifications which are correlated with the acoustic and/or articulatory realization of the phoneme. Linguistically, these features serve to describe “natural classes” of sounds which tend to pattern together with respect to phonological rules. The primary objective of this study is to examine the relation of these linguistic features to the mental representation of speech sounds through a series of speech perception experiments.

The role of linguistic features in speech perception has been under investigation for more than forty years. A variety of experimental procedures has been used, including the detection of mispronunciations, similarity scaling, dichotic listening, and consonant confusions. Despite the inherent difficulty of isolating the features (due to the multiple feature specifications of each phoneme and the variation in acoustic cues for each feature), these studies have been successful in revealing a listener awareness and utilization of features in speech perception tasks. Early studies in the detection of mispronunciations and similarity scaling of speech sounds demonstrate that listeners are sensitive to the number and nature of features shared among consonants and thus the psychological reality of features is affirmed. Dichotic listening experiments lend strong support to the hypothesis that features are extracted and suggest that they are processed independently. Further support comes from consonant recognition and consonant confusion studies. In particular, consistency in confusion patterns reflects the sorting of

consonants by features and suggests that certain features, such as place of articulation, are used to differentiate within more prominent feature groupings such as manner or voicing.

A review of these studies, as well as three that suggest a nonindependence of feature processing, illustrates the need to determine more precisely the nature of feature interactions, and motivates the experiments presented here. Although the involvement of features in consonant and word recognition has been confirmed, many basic questions remain. Specifically, how are these features organized with respect to the mental representation of the speech sounds? Do the features contribute equally to the phoneme representation, or are there dependencies among features? What is the nature of feature processing? Are certain features more easily or more accurately perceived? Is there a hierarchy of feature processing?

1.1 Detection of mispronunciations

Cole (1973) investigated the psychological reality of features using a mispronunciation detection task. Forty-five subjects were asked to detect mispronunciations in a recording of passages of Lewis Carroll's *Through the Looking Glass*. The mispronunciations were produced by changing one consonant sound in three-syllable words by one, two, or four features. There were forty-five of these stimuli, with 10 additional monosyllabic mispronunciations used as distracters. One-third of the mispronunciations occurred in the onset of the first syllable, while the other two-thirds were equally divided into the coda position of the second and third syllables.

The results showed that linguistic features are involved in word recognition. Only 30% of the one-feature changes were detected, while 60% and 70% of the two- and three-feature changes were detected. Mispronunciations were detected equally often in the three syllable positions, but reaction times were much longer for stimuli with first-syllable changes, presumably because more information is needed after just the first syllable to recognize the word. The results provide evidence for the psychological reality of features and suggest a correlation between the number of feature differences and the magnitude of perceptual difference.

Marslen-Wilson and Welsh (1978) continued along the same lines investigating perception of mispronunciations with both a shadowing task and a detection task. For the shadowing task, subjects were asked to repeat the material as they heard it, and were not told that mispronunciations would be present. In the detection task, the subjects were given a transcript and asked to circle relevant portions whenever they heard mispronunciations in the recorded version. The stimuli were similar to those used in Cole's study, with mispronunciations varying from one to three feature changes,

embedded in a normal prose passage. The experimental design contained two forms of constraint: a) the position of the changed consonant in the word (first or third syllable), and b) the predictability of the word in context (high or low).

For the shadowing task, reactions to the mispronunciations were fluent or nonfluent restorations, or exact or incorrect repetitions. Almost half of the responses (49%) were fluent restorations, that is, the subjects restored the mispronounced word without hesitation or stumbling. Exact repetitions accounted for 37% of the responses, and less than 15% were incorrect repetitions or nonfluent restorations. Of the fluent restorations, the overwhelming majority, 74%, was for stimuli that had only one feature change.

Results for the detection task were consistent with those of the shadowing task. The miss-rate for one-feature changes was 67%, a result which is not statistically different from the restoration rate above. For the three feature changes the miss-rate was only 6%. The contextual constraint had no significant effect in either experiment. The syllable constraint did have an effect in that miss-rates and fluent restorations increased when the change was in the third syllable, but the experimenters point out that this effect could have been influenced by the fact that the third syllables never got primary stress, as many of the first syllables did. Thus, the third syllables may have been less perceptually salient to the subjects.

In general, the two studies showed that listeners were least sensitive to mispronunciations that involve only one feature change. This outcome implies that the featural specifications of phonemes are more than just a useful linguistic description since the process of word recognition is affected differentially depending upon the number of shared features.

1.2 Scaling studies

Magnitude estimation (or scaling) studies are commonly used in the field of psychophysics to investigate perceptual scaling of qualities such as loudness, brightness, and similarity. In such work, subjects are asked to rate some quality of a stimulus on a scale. Results are then used to determine which characteristics of a stimulus contribute to a perceptual judgment. Scaling studies in speech perception have shown that the perceptual distance between speech sounds can be explained in terms of feature differences since the perceived similarity of two speech sounds depends upon the number and nature of shared features.

Peters (1963) was among the first to investigate the perceptual similarity of speech sounds using a scaling task. Subjects were asked to judge the perceptual similarity of natural speech tokens they had previously recorded. Thus, subjects made judgments about

their own speech sounds. The results for each subject were depicted in a two-dimensional consonant space, the structure of which was interpreted on the basis of articulatory features. Although subjects showed variation in the particular features used to group the consonants, the most important factor was manner, with voicing or place represented as within-group distinctions. The consonants for a particular subject, for example, were sorted according to the following manner distinctions: sibilants and fricatives, affricates, plosives, nasals, and glides. The second dimension for this subject was voicing for sibilants and fricatives, and place for stops. Despite evident variability among subjects, the results showed a clear tendency for all subjects to group by manner features. This outcome suggests that manner may play a more central role in the mental representation of consonants than place or voicing.

Greenberg and Jenkins (1964) used magnitude estimation in their investigation of similarity judgments of speech sounds. The study involved paired comparisons of English voiced and voiceless stop consonants followed by the vowel [a], with three different experimental designs: i) aural plus visual stimuli; magnitude estimation task; ii) aural stimuli only; magnitude estimation task; and iii) aural stimuli; rating scale task. For the magnitude estimation task, subjects were instructed to listen to the first pair of stimuli and to assign a number indicating the degree of similarity they felt the pair to have. For the next pair of stimuli, they were instructed to base their similarity judgment on the scale of their first judgment (i.e., if the stimuli were considered to be twice as different, the first rating should be doubled). The subjects given the rating scale task were provided with a six-point scale that ranged from “extremely similar” to “not similar”.

The results were consistent across subjects and experiments. That is, the visual information made no difference, nor did the estimation vs. rating scale. The rankings were averaged, and the order from most similar to least was as follows: bd, gd, pb, kg, bg, pt, kt, td, pd, pk, tb, tg, kd, kb, pg. The relative importance of the presence and absence of features for voicing and place of articulation were measured and three main conclusions were drawn. First, sounds differing by only one feature were consistently judged more similar than sounds differing by two features. Next, the distances between places of articulation were ranked in the following ascending order: labial-alveolar, alveolar-velar, labial-velar. Third, agreement in the presence of voicing was a much greater factor in judgments of similarity than agreement in the absence of voicing, suggesting a markedness of features.

Mohr and Wang (1968) expanded the scope of the previous experiment by including more subjects and more stimuli. Their four different sets of stimuli consisted of: a) voiced

and voiceless stops and nasals; b) labial and alveolar stops and fricatives; c) oral and nasal vowels; and d) front unrounded and rounded vowels and back vowels.

The similarity judgments were again consistent for subjects across experimental conditions. The results were examined in terms of features and it was found that opposite specification for certain features had significant effects on similarity scores. The features for which the effect of opposite specification was most pronounced ranked as follows: nasal, voiced, continuant, labial, alveolar, and velar. Thus two stimuli that shared the feature [+nasal] were judged to be more similar than two stimuli that did not share this feature, and a stimulus pair which shared nasality was judged to be more similar than a pair which shared voicing. The ranking of features is consistent with previous studies.

In summary, the scaling studies show that similarity judgments for speech sounds are directly correlated with feature sharing. The more features two stimuli share, the more similar they are perceived to be. Moreover, certain features seem to carry more perceptual weight than others for judgments of similarity, a finding which suggests that certain features may be more salient to the perceptual system.

1.3 Dichotic listening studies

Dichotic listening studies present competing stimuli simultaneously to subjects binaurally. This is one way of taxing the perceptual system without any degradation of the stimuli. Dichotic studies allow researchers to investigate feature processing in speech perception by examining the confusion patterns from consonant identification tasks that present stimuli that differ by one or more features.

One of the first studies to investigate feature processing using a dichotic presentation was Studdert-Kennedy and Shankweiler (1970). They studied dichotic effects in the perception of initial and final stop consonants (voiced and voiceless). Subjects were binaurally presented with competing nonsense CVC stimuli and asked to identify both stimuli. A right ear advantage for consonant sounds was found for all subjects.

The most important finding from this experiment was that when a stimulus pair shared a feature (voicing or place), errors were less likely than when the pair had a double contrast. Also, performance varied depending upon which feature was shared: shared place resulted in fewer errors than shared voicing. That is, when the pair contrasted in voicing but shared place, there were fewer errors in identification than when the pair contrasted in place but shared voicing. The researchers concluded that the features are extracted independently. This was supported by an error analysis of the percentage of features correct on single error responses of double contrast pairs. Voicing was correctly identified in 72% of these responses, while correct place identification accounted for only

19% of the responses. In other words, for the double contrast pairs, voicing was more often accurately identified than place. Partial explanation for this result could stem from the lack of a right ear advantage to tones. Voicing corresponds to the fundamental frequency, or pitch, of speech. Perhaps the lack of a right ear advantage is due to a lower, more peripheral processing of this information.

Pisoni and McNabb (1974) investigated the feature-sharing effect in a backward masking paradigm. They used synthetic CV stimuli consisting of voiced and voiceless stop consonants, which were presented with a short lag in onset times. The target and masking stimuli never shared place (targets were labial or alveolar, maskers were velar) but they shared voicing in half of the trials. The results replicated the previous findings in that performance was higher for pairs that shared voicing.

In an effort to determine if manner features are independently extracted and processed in the context of competing stimuli Blumstein, Tartter, and Michel (Blumstein 1974) investigated the dichotic perception of stops, nasals and fricatives. The stimuli were CV syllables differing by a one feature contrast of nasality (ba-ma, da-na) or continuancy (ba-va, da-za), or by a two feature contrast of nasality and place (ma-za, na-va) or continuancy and place (ba-za, da-va). The researchers were interested in the following types of evidence for feature extraction: a) a right ear advantage for nasals and continuants; b) the presence of interaction between competing stimuli based on feature sharing; and c) the occurrence of blend errors.

The results showed a right ear advantage (REA) for fricatives and stops, and a non-significant REA for nasals. The percent correct identification was significantly higher for fricatives which shared place, and non-significantly higher for nasals sharing place. There were very few blend errors for either the nasals or the fricatives, although the overall performance for nasals was extremely high, so interpretation of the errors is difficult. The researchers concluded that manner features might be extracted at a lower processing level than the features of place and voice.

A more extensive study by Hayden, Kirstein and Singh (1979) examined the role of distinctive features in the identification of 21 dichotically presented, syllable initial consonants. They used natural speech tokens of all English consonants except [ŋ], which does not occur syllable initially in English. The stimuli were edited to 377.5 ms, but no attempt was made to match consonant duration per se. The subjects were asked to attend to one ear at a time and to identify only the stimulus presented to that ear.

The results for stops were consistent with previous findings: a large REA, many blend errors, and increased identification when the competing stimuli contrasted by only one feature. The results for the other consonants differed in that the percent correct

identification increased as the number of feature differences increased. Intrusion errors were greater when the stimuli differed by fewer features and this was consistent across all manner class comparisons. Blend errors were most frequent for stops competing with stops but were significantly above chance for continuants competing with continuants and for stops competing with continuants, affricates, and nasals. Also of interest, the magnitude of the right ear advantage was found to depend on manner (stops and nasals had significant REA) and not on the number of feature differences.

The results of this study, when taken with the previous studies, suggest that identification of two stimuli is more accurate when those two stimuli share features, but that accuracy in differentiating two stimuli and identifying just one of them improves when they differ by more than one feature. So when subjects are asked to attend to only one ear, correct identification improves when the competing stimulus is very different, and worsens when the competing stimulus is more similar.

Dichotic listening studies have contributed greatly to the evidence that consonants are processed based on their distinctive features. The occurrence of blend errors, in particular, provides compelling evidence that features are extracted independently. In a true double blend error, the subject has in essence extracted all of the correct features and then reassigned them to the consonants incorrectly.

1.4 Consonant recognition/confusion studies

Miller and Nicely (1955) studied consonant identification at various signal to noise ratios combined with no filtering, low-pass filtering, and high-pass filtering conditions. The stimuli consisted of 16 English consonants, [p, t, k, f, θ, s, ʃ, b, d, g, v, ð, z, ʒ, m, and n], followed by the vowel [a]. Subjects were asked to identify all stimuli as best they could, and the results were confusion matrices. In the high-pass filtering condition, errors were scattered randomly, but consistencies were found for the other conditions. Errors were analyzed in terms of correct identification in terms of the following features: voicing, nasality, affrication, duration, and place of articulation. Percent relative information transmitted for each of these features was also calculated for each experimental condition. Results of these analyses show that voicing and nasality information is least affected and place of articulation information is most affected by low-pass and noise conditions. The consonant confusion patterns and relative information transmission data also suggest that features are processed independently of each other since the information transmitted for each feature roughly adds up to the total information available in the stimuli.

Wish and Carroll (1974) performed an INDSCAL analysis on the entire set of Miller and Nicely confusion matrices. This analysis involves determining the stimulus coordinates and the dimension weights for the subjects that account for the variation in the matrices. Thus the INDSCAL analysis output specifies the dimensions along which the data pattern. The advantages of this analysis are twofold. Firstly, the stimulus configuration included dimensions relevant to all experimental conditions. Secondly, the dimension weights provided information about the saliency of different dimensions under different kinds and degrees of degradation.

The results consisted of six interpretable dimensions along which the confusion data could be plotted: 1. voicing (voiced vs. voiceless) vs. 2. nasality; 3. sibilant frequency vs. 4. sibilance; 5. direction of F2 transition (rising vs. falling) vs. 6. voiceless stops vs. voiceless fricatives. For each of these dimensions, the relative dependency on the particular acoustical conditions was given. Thus the importance of 1 and 2 increased with increasing degrees of low-pass filtering or noise, and decreased with high-pass filtering. This is not surprising since the acoustic correlates of voicing and nasality involve low frequency information. The important conclusion from this reanalysis of the Miller and Nicely data is that the dimensions along which the data pattern correspond to the feature composition of the stimuli. This is further support for feature extraction and relative independence of feature processing.

In an experiment similar to that done by Miller and Nicely, Dubno and Levitt (1981) studied confusion matrices from syllable identification in quiet and noise. The principal difference between this study and Miller and Nicely (1955) was the type of stimuli used, namely syllables representing nearly all the CV and VC syllables which can be formed with English consonants combined with the vowels /a, i, u/. Subsets of the stimuli included either voiced or voiceless consonants, but no voicing contrasts were present within the sets since the developers noted that voicing errors are infrequent. The subjects identified the syllables by means of a response set that included nearly all of the most common consonant confusions as possible choices.

Overall performance was significantly better for voiced consonants than for voiceless consonants, with the largest differences in percent correct identification in the very lowest and highest speech levels (dB SPL). Initial consonants were much more accurately identified than final consonants at every signal level, and consonants in the context of /a/ were more accurately identified than those in the other vowel contexts.

An analysis of the consonant confusions included target/response pairs with greater than chance percentages of occurrence. Confusions were categorized in terms of the type of error made. Results show that place errors were much more frequent in both quiet and

noise than manner errors or place/manner errors. The large number of place errors for stimuli in noise is consistent with the findings of Miller and Nicely.

Contrary to the results of Miller and Nicely, though, was performance on nasals in noise. Nasals were the most difficult to identify in noise in Dubno and Levitt's study, while Miller and Nicely reported almost perfect identification of these consonants. One explanation for this is the fact that Miller and Nicely used white noise as a masker, but Dubno and Levitt's cafeteria noise contained additional low frequency energy that would more likely mask the nasals. As was found in the Wish and Carroll (1974) study above, there is a clear and not surprising dependency of the dimension (nasality) on the particular acoustical condition of the experiment.

The same Nonsense Syllable Test used in Dubno and Levitt's experiment was used by Gelfand, Piper, and Silman (1985) to investigate consonant recognition in quiet as a function of aging among normal hearing subjects between the ages of 20 and 65. The effect of initial vs. final position and vowel context was consistent with the previous study. Percentages of information transmitted were calculated for the confusion matrices taken together and for consonant features appropriate for each matrix. For the youngest subjects, nasality, sibilance, and glide/liquid quality information were transmitted most precisely, whereas stop, frication, and place information were transmitted less effectively. The percentage of information transmitted for these last three features decreases significantly as the age group of the subjects increases. The most common consonant confusions for all subjects were between /f/ and /θ/ and between their voiced correlates, /v/ and /ð/. Interestingly, the direction of the confusions that depended on position and vowel context were also identical across subjects and age groups.

In general, results from consonant recognition and consonant confusion studies are further evidence for the involvement of feature processing in the perception of speech sounds. The consistency of confusion patterns suggests predictability based on feature interactions. This implies the existence of interdependency among features, despite the evidence from dichotic listening and the information transmission data that suggest that the processing of features is independent.

1.5 Interdependence of feature processing

Results from a few studies make the counter-suggestion that feature processing is in fact interdependent. Miller (1977) used synthetic speech in a categorical perception experiment to show that perceived place is contingent on perceived voicing and vice versa. By systematically manipulating the timing relations that differentiate voiced from voiceless consonants (VOT) for stimuli which differed only in the direction and extent of

the second and third formant transitions (the acoustic information associated with place distinctions) identification of voicing varied depending upon the place cues in the signal. In a second experiment, by systematically manipulating the spectral properties (the rate, direction and extent of formant transitions), identification of place (labial vs. alveolar) was shown to vary depending upon the voicing or nasality cues in the signal. The results clearly show an interaction of feature processing.

Carden, Levitt, Jusczyk, and Walley (1981) also found that perceived place of articulation was dependent on perceived manner. This result comes from an experiment that presented the same acoustic stimuli (truncated natural syllables) under different manner labels to subjects in a forced choice identification task. The researchers found that the place feature was dependent on the manner value that the listener assigned to the stimulus rather than on any acoustic cues to manner.

In a consonant recognition study using forward and backward gated VCV stimuli, Smits (1998) found that the temporal distribution of perceptually relevant information differs for manner, place, and voicing. For the consonants in VCV stimuli, manner information seems to be available prior to place information, which is available prior to voicing information in the duration of the stimuli. In addition, voicing information is available at an earlier point for fricatives compared to stops. The author concludes that in consonant recognition, manner is classified first, and that a determination of where to extract information about place and voicing depends on the result of the manner classification. This implies an interaction of features such that classification of place and voicing is dependent to some extent on manner information.

These studies are intriguing in that they support a dependency between features. One important variable that the first two studies altered, however, was duration. It is possible that the place, manner, and voicing features perceived in these studies were also crucially dependent on the duration of the stimulus. Certainly Smits' study reveals that the amount of featural information available to the subject varies depending upon the gate of the stimulus presented, which is directly correlated with duration.

The studies reviewed here all give strong evidence for the existence and use of linguistic features in the perception of speech sounds. Although measurable progress has been made in determining which features are relevant to consonant recognition, the exact nature of the interaction of these features is still unknown. With this in mind, two experiments were designed to investigate a possible hierarchy of feature processing and the interaction of features in the identification of English consonants.

The purpose of the present study is to explore any possible feature interactions involved in consonant identification. Of interest is the relative perceptual saliency of the

consonants and of the linguistic features which can categorize them. It is possible that some hierarchy of difficulty exists, that is, that perhaps certain consonants or certain linguistic features are more easily or more accurately identified than others. It is also possible that feature processing is interdependent, such that certain feature information facilitates consonant identification or the perception of other features. In order to examine these possibilities, the present experiments were designed to elicit consonant confusions by gating the consonant portion of the stimuli presented to listeners in an identification task. This study attempts to characterize more precisely the nature of feature interactions by analyzing the resulting consonant confusions in terms of the features that were correctly perceived by the listeners.

2 Pilot experiments

The results of a previous gating experiment on the role of features in the identification of English consonants led to a modification of the stimuli to be used in the present experiment. When increasingly shorter consonant tokens were presented in isolation, there was a strong tendency for listeners to identify them as stop consonants. This judgment most likely reflected the fact that the absolute duration of the stimuli was so short. In order to circumvent this problem, it was necessary to mask the duration of the stimuli without masking the other acoustic information in the stimuli. After investigating the use of noise and babble in speech perception studies, a subset of stimuli was created in which consonant tokens with relatively short and long duration conditions were either embedded in noise or babble, or had noise or babble appended to them. These stimuli were presented to subjects in an identification task

2.1 Modifications suggested by noise and babble used in speech perception studies

One of the most relevant studies using noise is that of Miller and Nicely (1955). They found that identification of consonants embedded in white noise at a signal to noise (S/N) ratio of -6 dB resulted in a non-random confusion matrix. Voicing and manner features were much less affected by the noise than place features in this study, and Miller and Nicely point out that the noise effectively masks the spectral characteristics of the consonants which inherently contain noise, i.e. frication and aspiration. Replicating this aspect of Miller and Nicely's study, however, could provide interesting results in the current gating experiment, since this study presents the consonant tokens excised from their original CV environment, whereas Miller and Nicely embedded entire syllables. In addition, the current study advances this line of research through the presentation of

mutually exclusive subsets of the stimuli in Experiment 2. Therefore, embedding the consonant tokens in noise was one manipulation considered in this pilot study.

The notion of appending noise to the consonant tokens was suggested by Smits (personal communication) and is supported by a study by Pols and Schouten (1978). Pols and Schouten found that replacing deleted initial consonants with noise eliminates the abrupt onset caused by the incomplete syllable and improves identification of the deleted consonant. For the present study, it was hoped that appending noise would eliminate the abrupt truncation of the consonant and decrease the subjects' tendency to judge the consonants in the shorter duration conditions as stop consonants.

Crandell (1992) used multi-talker speech babble as the noise competition in a study investigating speech-recognition susceptibility to noise in elderly listeners. This babble was derived from the SPIN test, created by Kalikow et al. (1977). Kalikow et al. generated babble noise by mixing 12 sound files consisting of two repetitions each of six speakers (three males and three females) reading continuous text. These studies using speech babble prompted the decision to create such babble and investigate the effect of using it in place of the Gaussian noise in the creation of stimuli (i.e., embedding the consonants in babble or appending babble to the consonant tokens).

2.2 Pilot experiment 1

2.2.1 Methods

The recorded stimuli were CV utterances consisting of the consonants [p, t, f, s, b, d, v, z,] followed by the vowel [a]. The stimuli were produced by a male native speaker of American English. The speaker was recorded in a soundproof booth, using a cardioid microphone (Electrovoice, model RE 20) and a high-quality cassette recorder (Carver, model TD-1700). The speaker produced three repetitions of each target syllable in a carrier phrase (Say ____ to me) at a normal speaking rate. The recordings were then digitized on a SUN SPARCstation2 at 22050 Hz and stored as files to be processed by the commercial software package WAVES+/ESPS (Entropic, Inc.). For the present experiment, one out of the three repetitions of each target syllable was selected on the basis of its sound quality and segmentability. Sound quality refers to the informal judgment by a trained phonetician that a given token was representative of a particular consonant, and segmentability refers to the presence of clear onset and offset cues in the waveform and/or spectrogram analysis.

The tokens were edited directly on a graphics display terminal, using simultaneous waveform and spectrogram displays. Measurements and segmentations were made at the zero crossings of the waveform. For stops, consonant onset was defined as the point of

release after silence in the waveform (or just a voice bar in the spectrogram for voiced stops). For fricatives, onset was defined as the onset of aperiodic noise in the waveform, or the presence of high frequency energy in the spectrogram. Consonant offset was defined as the point at which the transitions leveled off and the steady state portion of the following vowel began. The consonant portion of each target syllable was excised and used in its entirety in creating the longest stimuli (the 100% condition stimuli).

This “phone-plus-transition” segmentation procedure, based on Hertz (1991) was chosen to ensure that the excised consonant tokens contained comparable information. Since transition information is included in the aspiration of the voiceless stops, it can be argued that inclusion of the transitions in the voiced stops is warranted. Furthermore, it was determined that using a more traditional definition of consonant offset, such as the onset of periodicity of the following vowel, would be problematic in the creation of the shorter stimuli. The durations of the voiced stops, which would then include solely the burst, would be too short to manipulate in the gating procedure (a 10- to 15-millisecond burst cannot reasonably be gated to include just 10 percent of its original duration). Thus for the voiced stops it was necessary to include the transitions into the following vowel. In order to isolate the effect of including this vowel information, it was decided that the inclusion of the following vowel transitions for all of the consonants in the 100% condition was necessary. If identification improved significantly between the 50% and the 100% conditions for the fricatives and voiceless stops, and if identification of the voiced stops was significantly better than identification of the other consonants at the shorter conditions, then this could be attributed to the contribution of vowel transition information. If, on the other hand, correct identification of the voiced stops was found to be comparable to the correct identification of the other consonants at each condition, then the suitability of the phone-and-transition segmentation procedure for these experiments is confirmed.

Preserving the relative durational differences of the different consonants, the excised consonants were then cut back from offset to 50%, 40%, 30%, 20%, and 10% segments of their original duration, each preserving the consonant onset. These will be referred to as the 50% to 10% conditions. The 60% to 90% conditions were omitted from the experiment in light of previous studies which show that listeners can accurately identify certain features of phonemes with absolute durations shorter than or roughly corresponding to the 50% condition for the consonants in this study. Blumstein and Stevens (1980), for example, showed that listeners could reliably identify the place of articulation of stop consonant-vowel syllable segments as short as 10 to 20 ms. Jongman

(1989) found that the frication duration required for identification of the fricatives included in this experiment ranges from 30 to 50 ms.

The first pilot stimuli were created from a subset of the consonant tokens to be used in the experiment, namely tokens from the 10% and the 50% condition for each of the eight consonants, [p, t, b, d, f, s, v, z]. These 16 consonant tokens were either embedded in noise, appended with noise, embedded in babble, or appended with babble. Thus there were a total of 64 pilot stimuli consisting of 32 noise stimuli and 32 babble stimuli.

For the noise stimuli, a 250 ms token of Gaussian noise was generated on a SUN SPARCstation2 at a sampling rate of 22050 Hz. This noise was used to create stimuli embedded in noise and appended with noise.

For the embedded stimuli, the 250 ms noise file, which was longer than any of the consonant tokens, was added to the consonant files such that the noise always began 25 ms prior to the consonant onset. If the consonant token had been centered in the noise, the token duration could have been evident from the time differences between noise onset and consonant onset for the different stimuli. On the basis of the assumption that listeners are more sensitive to consonant onsets than offsets, consonant onset within the noise was kept constant.

For the appended stimuli, a portion of the noise file was appended at the end of the consonant file such that the total stimulus duration was 250 ms (including 25 ms of silence before consonant onset). Thus the duration of noise varied depending upon the duration of the consonant.

In the embedded condition, the original RMS power of the noise (54 dB) was altered such that the S/N ratio for each stimulus was -6 dB. This ratio was chosen on the basis of Miller and Nicely's (1955) results which showed decreased identification yet non-random consonant confusions at this ratio. The power of the noise was 6 dB more than the power measurements taken for each of the consonants, in each of the conditions. Power measurements for the consonant tokens ranged from 53 dB for the 10% condition of [f] to 74 dB for the 100% condition of [b].

For the appended condition, the power of the noise was matched to that of the consonant to facilitate the streaming integration effect and thus make listeners less sensitive to the absolute duration of the consonant token.

Multi-talker speech babble was created following the method used by Kalikow et al. (1977). Four speakers (two males and two females) were recorded reading continuous text. After each speaker's recorded passage was digitized onto a SUN SPARCstation2 at 22050 Hz, three different 2 second samples were taken and stored in separate speech files. All twelve files were then digitally added, resulting in 2 seconds of babble. The

babble sounded like uninterpretable speech in which phonemes and possibly syllables could be picked out if one attended closely. When the consonant tokens were embedded in this babble, even the longest tokens were judged to be completely unidentifiable to the experimenter. The appended stimuli were equally difficult to identify, and so the use of this babble was rejected. As other consonant sounds in the babble were hypothesized to be masking the stimuli, vowel babble was created as an alternative.

Since vowel babble would not contain the exact same acoustic correlates as the consonant tokens (e.g., bursts, aperiodic energy in the higher frequencies, or rapid transitions), it was hypothesized that vowel babble would not hamper identification as much as the previous speech babble, but would still mask the duration of the stimuli more effectively than noise. This is assuming that subjects might be able to integrate the streams of the vowel babble and the consonant token and hear them as coming from the same source. The vowel babble was created in much the same way as the speech babble. Six speakers (three male and three female) produced each of 11 English vowels, [i, I, e, ε, Λ, æ, u, o, ɔ, a, ɑ] twice in isolation, deliberately lengthening the vowels to ensure at least 250 ms of steady-state formants. The recordings were digitized and a 250 ms segment was excised from the center of each vowel. All vowel files were added, producing an atypical- sounding [a] in which spectrographic analysis revealed many formant frequencies.

Adding the vowel babble to the consonant tokens was done in a similar manner as described above for the noise stimuli. For the embedded stimuli, the 250 ms vowel babble was added to the consonant file such that the babble began 25 ms prior to consonant onset. For the appended stimuli, a portion of vowel babble was appended to the consonant file such that the total stimulus duration was 250 ms (including 25 ms of silence before consonant onset). Thus the duration of vowel babble varied depending upon the duration of the consonant token.

As was the case for the appended noise stimuli, in both the embedded and appended vowel babble conditions, the RMS power of the babble was matched to the RMS power of the consonant token, which ranged from 53 dB to 74 dB. In all cases, the RMS power of the babble was decreased from its original 84 dB.

Two trained phoneticians participated in the pilot study. Subjects were asked to identify two repetitions of the stimuli presented randomly within each of the four experimental conditions: embedded in noise, noise appended, embedded in vowel babble, vowel babble appended. Subjects first identified two repetitions of the 50% condition stimuli in the two embedded conditions followed by the two appended conditions. On the

basis of their performance on these stimuli, subjects then identified both the 10% and 50% conditions of the babble-appended stimuli.

2.2.2 Results and Discussion

For the stimuli embedded in noise, identification was very poor, with only 45% and 27% correct identification for the two subjects. For the stimuli embedded in vowel babble, identification was slightly improved at 69% and 63%, and subjects reported little awareness of the absolute duration of the stimuli, but since neither subject reached criterion (80% correct) it was decided that embedding the consonant tokens hampered identification too much and this method was rejected.

While both subjects reached criterion performance in the noise-appended condition, with correct identification at 95% and 81%, subjects reported that the stimuli sounded composed of two very distinct elements - a consonant sound and a burst of noise. That is, despite the lack of silence between the consonant and the noise, it was felt that the absolute duration of the consonant portion was not masked by the noise. Thus, these noise-appended stimuli did not produce the desired effect. In the vowel-babble-appended condition, however, subjects reported hearing a strange-sounding syllable. Correct identification varied for the two listeners, at 100% and 69%, but both reported that the duration of the consonant was effectively masked by the babble.

These babble-appended stimuli were therefore promising enough to try in both duration-condition tests. Results for the 10% condition were very poor, with only 19% correct for each listener, while results for the second trial of the 50% condition decreased, with 88% and 63%. Although identification was very poor in the 10% condition, it is important to note that subjects did not show the tendency to judge these stimuli as stop consonants. This was taken as evidence of the effectiveness of appending vowel babble as a way to mask the absolute duration of the consonant tokens. The high identification scores for the 50% condition were taken as evidence that the appended babble does not interfere too much with the acoustic correlates of the consonant.

These results led to the decision to use these stimuli in the experiment, with just one final modification. Subjects reported that the amplitude of the consonants seemed unnaturally high in comparison to the babble, especially in the 10% condition. This was a result of having matched the RMS power of the babble to the consonant token. It was decided that a more natural approach would involve matching the power of the vowel babble to the power of the original vowel of the syllable from which the consonant had been excised. This final modification was subsequently made to the stimuli and a second pilot experiment was run.

2.3 Pilot experiment 2

2.3.1 Methods

The stimuli consisted of all of the consonant tokens to be used in the experiment (the 10%, 20%, 30%, 40%, 50%, and 100% conditions for [p, t, b, d, f, s, v, and z]) with vowel babble appended as in the first pilot study, but with a modification of the RMS power level of the babble. The power level of the babble was matched to the original power level of the vowel from which the consonant token had been excised. There were a total of 48 stimuli consisting of eight consonants at six gated conditions.

One trained phonetician and one monolingual American English speaker participated in the experiment. Subjects heard a random presentation of five repetitions of the stimuli, blocked by condition. Thus there were 40 stimuli per block and a total of 240 stimuli. The blocks of stimuli were presented in the following order: 40%, 30%, 10%, 50%, 20%, 100%. Before each condition, subjects were given 10 practice trials.

2.3.2 Results and Discussion

Table 1 shows the correct identification scores (in percentage) across all consonants for each condition for each subject.

| | Correct Identification (%) | |
|-----------|----------------------------|-----------|
| condition | Subject 1 | Subject 2 |
| 10% | 7 | 18 |
| 20% | 40 | 35 |
| 30% | 88 | 45 |
| 40% | 93 | 78 |
| 50% | 90 | 90 |
| 100% | 100 | 100 |

Table 1. Correct identification (in percent) for each condition.

These results suggest that correct identification improves linearly as the duration of the consonant portion of the stimuli increases. Scores are above chance (one in eight choices equals 12%) for both subjects as early as the 20% condition.

These pilot results were also analyzed by calculating correct identification of the features: voicing, place of articulation, and manner of articulation. Table 2 shows the correct identification of each of these features. For each feature, the scores (in percent correct) include responses that are consistent with the stimulus in terms of that feature alone and errors of the other two features are ignored.

| Condition | Correct Identification (%) | | | | | |
|-----------|----------------------------|-----------|---------------|-----------|----------------|-----------|
| | Voicing feature | | Place feature | | Manner feature | |
| | Subject 1 | Subject 2 | Subject 1 | Subject 2 | Subject 1 | Subject 2 |
| 10% | 53 | 68 | 58 | 60 | 53 | 55 |
| 20% | 53 | 75 | 88 | 80 | 88 | 80 |
| 30% | 95 | 73 | 95 | 83 | 90 | 70 |
| 40% | 93 | 83 | 100 | 98 | 98 | 98 |
| 50% | 90 | 98 | 100 | 93 | 100 | 100 |
| 100% | 100 | 100 | 100 | 100 | 100 | 100 |

Table 2. Correct identification of voicing, place, and manner.

Again, identification of each of the features improves as the duration of the consonant portion of the stimuli increases. These results suggest that feature information is indeed available to the subjects even in short conditions. Whether certain feature information is more easily or accurately perceived cannot be determined due to the limited scope of this pilot study, but these results were promising enough to motivate Experiments 1 and 2.

3 Experiment 1

The aim of Experiment 1 was to investigate the relative perceptual saliency of features in the identification of the English consonants [p, t, b, d, f, s, v, z] presented in six different gated conditions. A forced-choice task was used with the gating paradigm to elicit consonant confusions. In addition to an analysis of correct identification of the consonants for each condition, two types of analyses were performed on the consonant confusion data - correct identification of each feature and information transmission for each feature.

3.1 Methods

3.1.1 Stimuli

The stimuli to be used in the experiment were those created in the second pilot study. The stimuli consisted of the consonant tokens [p, t, b, d, f, s, v, z], excised from natural speech CV syllables and modified to include 10%, 20%, 30%, 40%, 50%, and 100% of their original duration, to which vowel babble was appended such that the duration of each stimulus was 250 ms (including 25 ms of silence before the consonant portion). Table 3 gives the durations of the consonant portions of the stimuli by condition.

| Condition | Consonant Durations (ms) | | | | | | | | mean |
|-----------|--------------------------|-----|-----|-----|-----|-----|-----|-----|------|
| | [p] | [t] | [b] | [d] | [f] | [s] | [v] | [z] | |
| 10% | 9 | 13 | 6 | 5 | 16 | 17 | 15 | 17 | 12 |
| 20% | 18 | 26 | 12 | 11 | 32 | 34 | 29 | 30 | 24 |
| 30% | 28 | 38 | 16 | 17 | 51 | 51 | 40 | 50 | 36 |
| 40% | 39 | 51 | 21 | 24 | 64 | 69 | 54 | 65 | 48 |
| 50% | 48 | 64 | 24 | 28 | 80 | 85 | 65 | 82 | 60 |
| 100% | 94 | 128 | 51 | 56 | 160 | 170 | 135 | 164 | 120 |

Table 3. Durations (in ms) for each consonant for each condition.

3.1.2 Subjects

Sixteen students at Cornell University voluntarily participated in the first perception experiment and subsequently received candy. All subjects were native speakers of American English who reported no history of speech or hearing disorders.

3.1.3 Procedures

The stimuli described above were transferred directly from the SUN SPARCstation2 to a Swan 386/25 PC. Using BLISS software (Mertus, 1989), Experiment 1 was designed such that five repetitions of each stimulus were presented to the subjects randomly within blocks of the different conditions. All stimuli were presented auditorily over Sony headphones with an inter-stimulus interval (ISI) of 3 seconds. Subjects were first presented with the block of stimuli of the 100% condition. Subjects then listened to the remaining blocks presented in a quasi-random order alternating longer and shorter duration blocks. For each condition, subjects had 10 practice trials, followed by 40 stimuli in the set.

The subjects were told that in each set (i.e., condition) they would hear syllables that consisted of one of the eight consonants on their answer sheets, as well as a strange-sounding vowel. The subjects' task was to identify the consonant token present in each stimulus, and indicate their choice by circling the consonant on their answer sheets. Subjects' performance on the 100% condition was used as the criterion for continued participation in the experiment. Two subjects were excluded for failure to meet a predetermined criterion of correctly identifying at least one token of each consonant and correctly identifying 85% of the stimuli in the 100% condition.

3.2 Analysis of results

The results for all fourteen subjects are compiled in six confusion matrices which are included in Appendix A. The following analyses were performed on the results: First, correct identification for each consonant for each condition was tabulated. Next, an analysis of correct identification of the linguistic features of voicing, place of articulation, and manner of articulation was done. For each analysis, ANOVAs were used to compare the mean identification scores for the different consonants and conditions. Finally, the covariance between stimulus and response for each condition in terms of the features above was measured using information transmission analysis.

3.2.1 Correct identification of consonants by condition

Figure 1 shows the correct identification scores across subjects for each consonant for each condition. A two-way analysis of variance (ANOVA) was conducted: Consonant by Condition. There was a main effect found for both consonant [$F(7,455) = 18.9, p < 0.001$] and condition [$F(5,455) = 214.0, p < 0.001$]. To examine any significant differences in the identification scores for the different consonants, a Newman-Keuls post-hoc test was performed. Of interest was whether certain consonants were more accurately perceived than others. The results showed that the mean scores for the voiceless stops, [p] and [t], and the voiced labial fricative [v] were each significantly different from the scores for the voiceless fricatives, [f] and [s]. In addition, the scores for [t] were also significantly different from those of the voiced stops, [b] and [d].

There was also a significant consonant X condition interaction effect [$F(35,455) = 5.9, p < 0.001$]. As shown in Figure 1, consonant identification scores improved as the duration increased. Scores for the voiceless stops, [p] and [t], and voiced fricatives, [v] and [z], increased rapidly and attained near perfection (above 90%) at the 30% condition. Scores for the voiced stops [b] and [d] improved slowly and showed a marked increase from the 50% to the 100% condition. Finally, scores for the voiceless fricatives, [f] and [s], rose steadily as the duration of the consonant increased.

As shown in Figure 1, there was no great increase in the identification scores of the fricatives and voiceless stops between the 50% and 100% conditions. The vowel transition information included in the 100% condition had no noticeable effect on the identification of these consonants. For the voiced stops, there was a large increase in the identification scores between the 50% and 100% conditions, but this cannot be due to the presence versus absence of vowel information since vowel transition information is included in 30% and 40% conditions for these consonants. It is important to note also that mean scores for the voiced stops differed significantly only from [t], which was the most

accurately identified consonant. These results indicate that the “phone-and-transition” segmentation procedure was appropriate for this experiment.

Newman-Keuls post-hoc tests for each of the consonants showed that the correct identification scores for the voiceless stops, [p] and [t], and the voiced fricatives, [v] and [z], increased significantly from the 10% to the 20% condition and from the 20% to the 30% condition. Scores for the voiced alveolar stop [d] increased significantly from the 10% to the 20% condition and scores for the voiceless alveolar fricative [s] increased significantly from the 20% to the 30% condition. The increase apparent in Figure 1 for voiced stops from the 50% to the 100% condition was not significant.

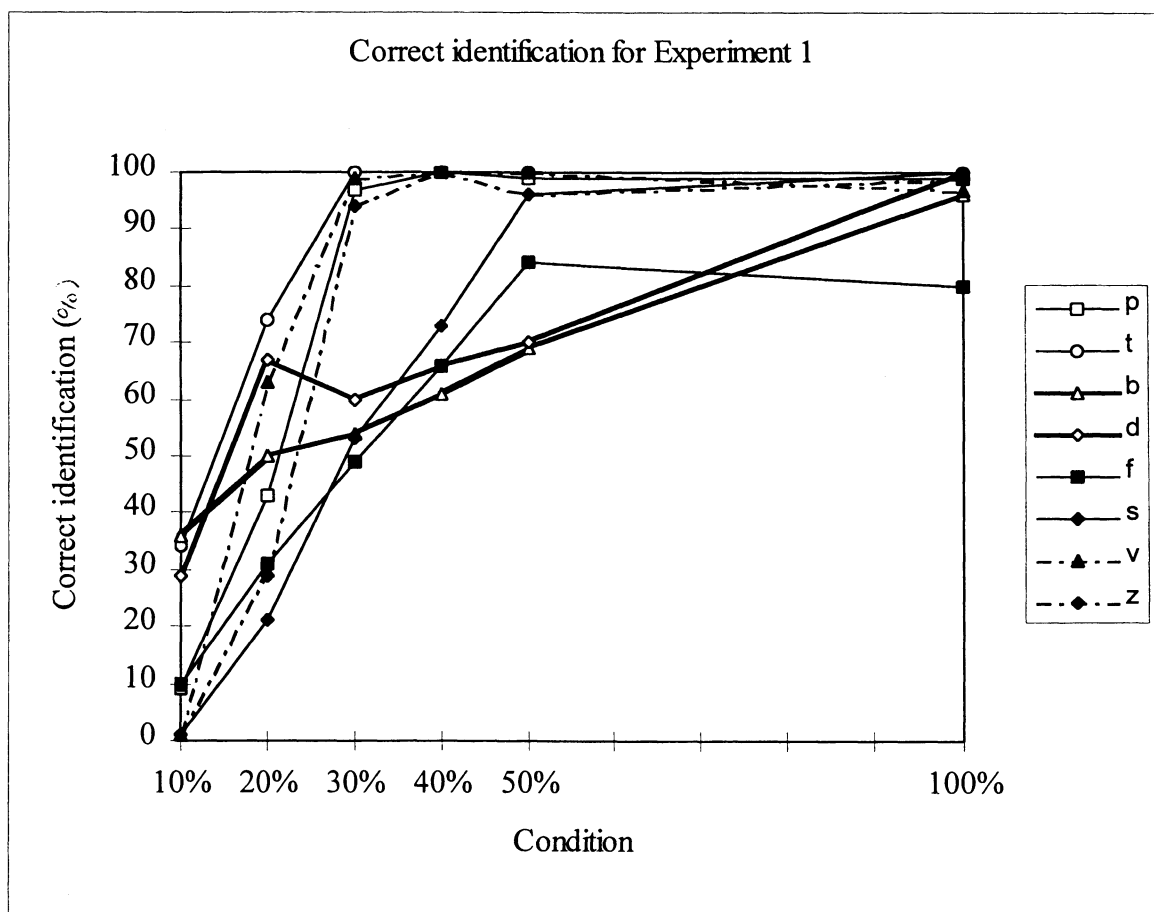


Figure 1. Correct identification (in percent) of each consonant at each condition for the Experiment 1.

3.2.2 Identification of voicing feature

Table 4 shows the identification scores, averaged over the fourteen subjects, for the identification of the voicing feature of the consonants by condition. The scores reflect correct identification in terms of voicing only, and errors of place and manner were disregarded. The consonants [p, t, f, s] were included in the voiceless category and [b, d, v, z] comprised the voiced category. In general, as the duration increased, correct identification of voicing also increased. A two-way ANOVA revealed a main effect for both consonant [$F(7,455) = 22.21, p < 0.001$] and condition [$F(5,455) = 64.67, p < 0.001$]. There was also a significant consonant X duration interaction [$F(35,455) = 7.19, p < 0.001$].

In general, perception of voicing was most accurate for the voiced consonants, and was above chance even for the 10% condition (although scores for [d] did not show the steady increase with increased duration that the other voiced consonants show). Of the voiceless consonants, voicing was most accurately perceived for the alveolars [t] and [s]. Although voicing errors for [p] were almost nonexistent by the 30% condition, errors for [f] persisted until the 50% condition.

A Newman-Keuls post-hoc test showed that voicing scores for [f] were significantly lower than those of all other consonants. In addition, the mean voicing scores for [d] and [p] were significantly lower than the scores for [t], [z], and [v].

| Condition | Correct Identification (%) | | | | | | | | Mean |
|-----------|----------------------------|-----|-----|-----|-----|-----|-----|-----|------|
| | [p] | [t] | [b] | [d] | [f] | [s] | [v] | [z] | |
| 10% | 16 | 46 | 77 | 80 | 31 | 60 | 87 | 80 | 60 |
| 20% | 47 | 84 | 71 | 77 | 44 | 86 | 83 | 84 | 72 |
| 30% | 99 | 100 | 83 | 60 | 54 | 76 | 100 | 99 | 84 |
| 40% | 100 | 100 | 99 | 69 | 71 | 81 | 100 | 100 | 90 |
| 50% | 100 | 100 | 91 | 70 | 94 | 97 | 100 | 100 | 94 |
| 100% | 100 | 100 | 100 | 100 | 96 | 100 | 100 | 99 | 99 |

Table 4. Correct identification (in percent) for the voicing feature of each consonant, for each condition.

3.2.3 Identification of place feature

Table 5 shows the correct identification scores, averaged over the 14 subjects, for the place feature of each consonant for each condition, regardless of identification of voicing

or manner. Categorization of the consonants was based on a distinction between a labial articulation (bilabial or labio-dental) and an alveolar articulation. The labials consisted of [p, b, f, v]. These were contrasted with the alveolars [t, d, s, z]. The table clearly shows that as duration increased, correct identification in terms of place of articulation also increased. A two-way ANOVA revealed a main effect for both consonant [$F(7,455) = 19.86$; $p < 0.001$] and duration [$F(5,455) = 52.22$; $p < 0.001$]. There was also a significant consonant X duration interaction [$F(35,455) = 5.72$; $p < 0.001$].

A Newman-Keuls post-hoc showed that place scores for [b], [f], and [z] were significantly lower than were place scores for the voiceless stops [p] and [t], and for the fricatives [s] and [v]. The scores for [b] were also significantly lower than those for [d].

| Condition | Correct Identification (%) | | | | | | | | Mean |
|-----------|----------------------------|-----|-----|-----|-----|-----|-----|-----|------|
| | [p] | [t] | [b] | [d] | [f] | [s] | [v] | [z] | |
| 10% | 80 | 81 | 59 | 39 | 49 | 99 | 83 | 40 | 66 |
| 20% | 93 | 91 | 70 | 84 | 79 | 77 | 91 | 49 | 79 |
| 30% | 99 | 100 | 73 | 100 | 87 | 86 | 99 | 96 | 93 |
| 40% | 100 | 100 | 63 | 97 | 94 | 93 | 100 | 100 | 93 |
| 50% | 99 | 100 | 77 | 100 | 90 | 99 | 100 | 96 | 95 |
| 100% | 99 | 100 | 96 | 100 | 83 | 100 | 97 | 100 | 97 |

Table 5. Correct identification (in percent) for the place feature of each consonant, for each condition.

3.2.4 Identification of manner feature

Table 6 shows correct identification in terms of manner of articulation, regardless of identification of voicing or place, for each condition, averaged over the fourteen subjects. The consonants were classified as stops ([p], [t], [b], and [d]) and fricatives ([f], [s], [v], and [z]). A two-way ANOVA revealed a main effect for both consonant [$F(7,455) = 33.89$; $p < 0.001$] and duration [$F(5,455) = 276.68$; $p < 0.001$]. There was also a significant consonant by duration interaction [$F(35, 455) = 19.66$; $p < 0.001$].

When a Newman-Keuls post-hoc test was performed to compare the mean scores of the different consonants, significant differences were found for the identification of manner for the stops and the fricatives. Manner scores for all four stops were significantly higher than those of all four fricatives, but at least for the shortest conditions this may

reflect a response bias such that the short fricatives are perceived as stops. Mean scores for the voiceless stops, [p] and [t], were higher than for any other consonants.

| Condition | Correct Identification (%) | | | | | | | | Mean |
|-----------|----------------------------|-----|-----|-----|-----|-----|-----|-----|------|
| | [p] | [t] | [b] | [d] | [f] | [s] | [v] | [z] | |
| 10% | 86 | 86 | 76 | 86 | 17 | 3 | 10 | 11 | 47 |
| 20% | 94 | 93 | 94 | 90 | 73 | 43 | 63 | 74 | 78 |
| 30% | 97 | 100 | 99 | 100 | 99 | 90 | 99 | 100 | 98 |
| 40% | 100 | 100 | 100 | 100 | 97 | 97 | 100 | 100 | 99 |
| 50% | 100 | 100 | 99 | 100 | 100 | 100 | 100 | 100 | 100 |
| 100% | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Table 6. Correct identification (in percent) in terms of manner, for each consonant, for each condition.

3.2.5 Comparative feature identification

One of the most interesting questions this study sought to answer is whether certain feature information was more (or less) easily or accurately perceived than other feature information. In order to answer this question, an ANOVA was performed to compare the previous findings of correct identification in terms of each of the features. That is, the information in tables 4, 5, and 6 was tested for an effect of feature. A main effect of feature was found [$F(2, 2013) = 4.05, p < 0.05$]. A Newman-Keuls post-hoc test showed that identification of voicing was found to be significantly poorer than identification of place or manner. This result shows that of the features included in this study, voicing was least accurately perceived.

3.2.6 Information transmission analysis

The examination of a possible response bias was achieved by measuring the covariance between stimulus and response. To this end, the data were subjected to information transmission analysis (Miller and Nicely, 1955). This analysis computes the amount of information transmitted by the stimuli in each condition, relative to the maximum amount of information available. Basically, information transmission takes into account the proportion of correct responses to each consonant, the frequency of each consonant in the stimulus, and the frequency of that consonant as a response. The equation for covariance of stimulus with response is:

$$T(x,y) = - \sum_{i,j} p_{ij} \frac{p_i p_j}{p_{ij}} ,$$

where $T(x,y)$ is transmitted information in bits per stimulus from an input x to an output y , p_i is the probability of an input stimulus i , p_j is the probability of an output response j , and p_{ij} is the probability of a joint occurrence of input i and output j . The maximum information available in the stimulus set is:

$$H(x) = - \sum_i p_i \log p_i .$$

Finally, the relative transmission is given by

$$T_{\text{rel}}(x,y) = T(x,y)/H(x).$$

Of interest is the comparison between relative information transmitted for each of the linguistic features of voicing, place of articulation, and manner of articulation. Figure 2 shows the relative information transmitted for consonant identification of each of these features for each condition. In general, the percentage of information transmitted for all three features increases as duration increases, with the largest increase for all features occurring between the 20% and 30% condition. Figure 2 shows that for Experiment 1, there is more information transmitted about manner than about place or voicing, except at the 10% condition, where place information is more salient. These results are consistent with the results of the previous analyses in that, overall, less information is transmitted about voicing than place or manner.

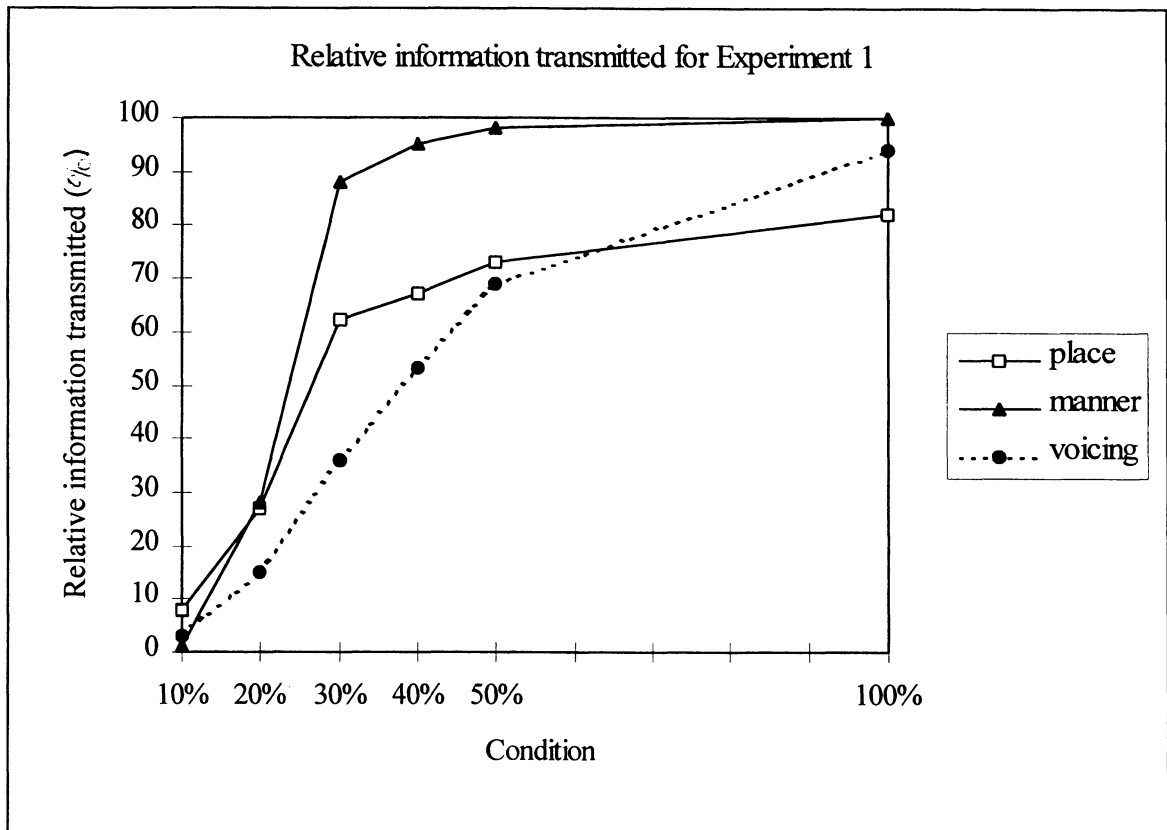


Figure 2. Relative information transmitted for voicing, place, and manner, for each condition.

3.3 Discussion

The aim of Experiment 1 was to investigate the nature of feature perception in the identification of English consonants. The first finding of interest was that there were significant differences between the identification scores for particular consonants, with the highest scores for the voiceless stops, [p] and [t], and the voiced fricative [v], and the lowest scores for the voiceless fricatives [f] and [s]. Next, the feature analyses showed that subjects were poorer at identifying voicing than place or manner. This finding was consistent with the information transmission analysis which showed less information transmitted about voicing than about place or manner at the 20%, 30%, and 40% conditions. One confusing finding was the discrepancy between the information transmission and the feature identification analyses concerning manner information. Despite the fact that more manner information was transmitted in the 30% to 100% conditions than place and voicing information, there was no statistical difference between the subjects' identification of place and manner.

The results from Experiment 1 illustrate the success of the methodology employed. The increase in identification scores corresponding to longer conditions confirms that the experiment design succeeded in eliciting consonant confusions in a systematic way that enabled meaningful comparisons across conditions, and the feature analyses showed that these confusions were not random errors. The consistency of the results of feature identification and information transmission analyses adds to the robustness of the findings.

4 Experiment 2

The goal of Experiment 2 was to investigate possible feature interaction in consonant recognition. Three groups of subjects identified the same eight stimuli from Experiment 1, but these stimuli were presented in two mutually exclusive sets of four stimuli which shared either voicing, place, or manner. Voicing is shared, for example, in the sets [p, t, f, s] and [b, d, v, z], and the group of subjects presented with these sets (the Voicing Group) had no decision to make about this feature in their identification of the consonants. The identification task remained the same as in Experiment 1, but as subjects identified all stimulus conditions from one set at a time, they had only four possible responses to choose from at any given time during the experiment. By manipulating the stimulus sets in this way, it was possible to examine the effect of given feature information on consonant identification. The results of Experiment 1 were used as a basis of comparison to reveal which feature information (i.e., voicing, place, or manner) facilitated consonant or feature identification.

4.1 Methods

4.1.1 Stimuli

Using the stimuli from Experiment 1 and the same BLISS software, six different stimulus sets were created. The experiment sets contained stimuli that shared either voicing, place, or manner. The voicing sets consisted of a voiced set [b, d, v, z], and voiceless set [p, t, f, s]. The place sets consisted of labials [p, b, f, v], and alveolars [t, d, s, z]. The manner set consisted of fricatives [f, s, v, z], and stops [p, t, b, d].

4.1.2 Subjects

Thirty-eight students at Cornell University voluntarily participated and received either candy or five dollars. All subjects were native speakers of American English who reported no history of speech or hearing disorders.

4.1.3 Procedures

The thirty-eight subjects comprised 3 groups, each of whom heard two experiment sets in which the same feature was shared. There were 13 subjects in the Voicing Group, 12 subjects in the Place Group, and 13 subjects in the Manner Group. Subjects in each group were thus given information about a particular feature in that they were not required to make a decision about that feature in their identification of the consonants. For each condition, five repetitions of the stimuli within each set were presented randomly to the subjects. For each experiment set, the blocks of stimuli were presented in a similar manner as in Experiment 1: subjects heard the 100% condition first, followed by a semi-random presentation of the other conditions. Before each block of 20 stimuli, subjects were presented with five practice trials. Subjects were told that they would hear syllables which contained one of the consonants labeled on their button boxes, and a strange-sounding vowel. For each stimulus, they were asked to use the index finger of their dominant hand to push the button labeled with the consonant they had most likely heard. The order of presentation of the two sets for a particular feature (e.g., voiced and voiceless) was randomized across subjects, as was the labeling of the consonants on the button boxes. Although the identification task and the instructions were similar to that used in Experiment 1, a crucial difference was the reduced number of stimuli in the presentation sets and the reduced number of possible response choices. In Experiment 1, subjects heard a random presentation of eight stimuli per block and had eight response choices while subjects in Experiment 2 heard a random presentation of four stimuli per block and had only four response choices at any given time.

As in Experiment 1, subjects' performance on the 100% condition was used as the criterion for inclusion in Experiment 2. Again, any subject who got less than 85% correct in identification of the consonants, or who misheard all tokens of a particular consonant in the 100% condition, was excluded from the results. This was the case for 7 subjects, 3 from the Voicing Group, 1 from the Place Group, and 3 from the Manner Group.

4.2 Analysis of results

Results for the three groups of subjects are compiled in confusion matrices included in Appendix A. The results are organized in the same way as the results for Experiment 1. After an analysis of correct identification scores for each consonant for each condition for each of the three groups, results were analyzed for correct identification of the two features relevant for each group (e.g. voicing analysis is not included for the Voicing Group since they had no voicing decision to make). For each analysis, ANOVAs were conducted on the responses from each subject group to compare the mean identification

scores for the different consonants and conditions. Finally, for each subject group, the results were subjected to information transmission analysis to measure the covariance between stimulus and response for each condition in terms of the relevant features.

4.2.1 Correct identification of consonants by condition for the three groups

Figure 3 shows the correct identification scores for each consonant across subjects in the Voicing Group. A two-way ANOVA revealed a main effect for both consonants [$F(7, 315) = 23.26, p < 0.001$] and condition [$F(5, 315) = 132.45, p < 0.001$], as well as a significant consonant X condition interaction [$F(35, 315) = 5.22, p < 0.001$]. A Newman-Keuls post-hoc test showed that identification of the voiceless stop [t] was significantly better than identification of the fricatives, [f], [s], and [z], and the voiced labial stop [b]. Identification of the voiceless alveolar fricative [s] and the voiced labial stop [b] was significantly worse than identification of all other consonants.

As shown in Figure 3, scores for all consonants increased with the duration condition. It is interesting to note the dramatic increase in identification between the 10% and 20% condition for all but the voiceless fricative [s]. Also striking is the high identification score for the voiceless alveolar stop [t] at the 10% condition. Scores for the voiceless fricatives, [f] and [s], and the voiced labial stop [b] differed from scores for all other consonants in that they do not reach perfection at the 30% condition.

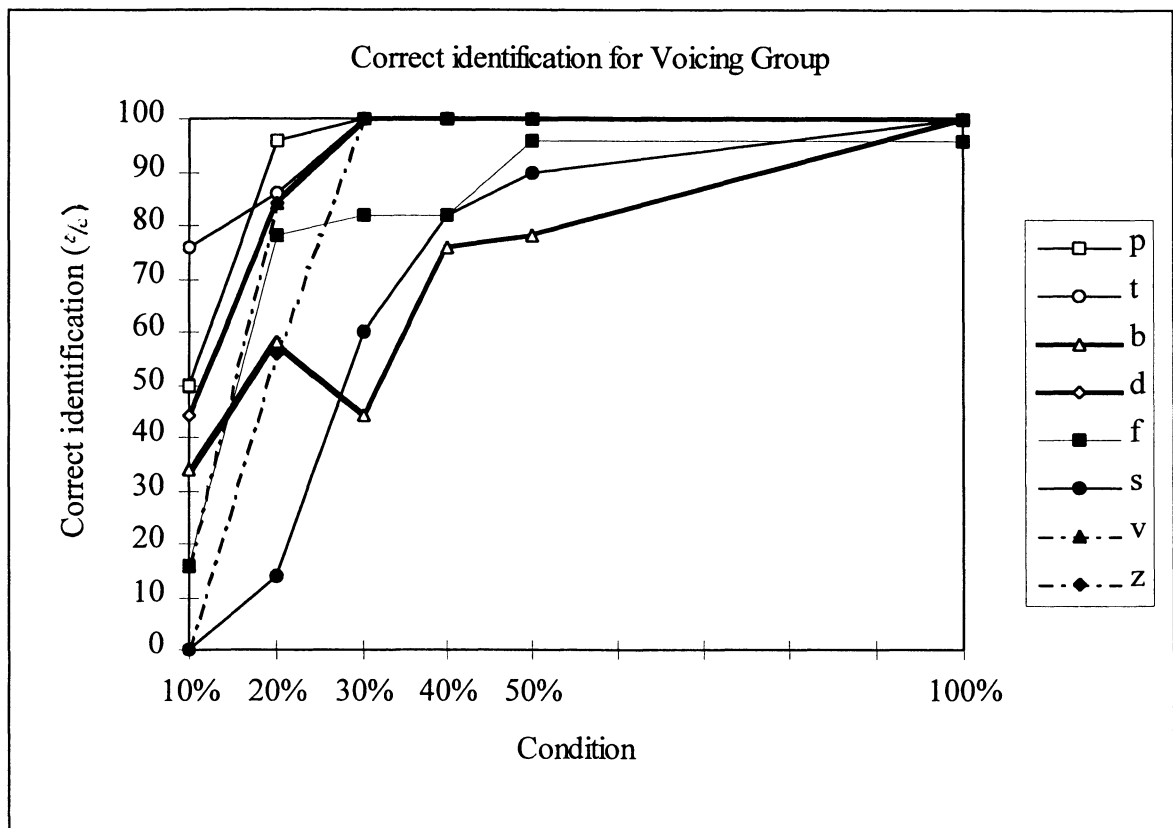


Figure 3. Correct identification for each consonant for each condition for the 10 subjects in the Voicing Group.

Figure 4 shows the correct identification scores for each consonant across subjects in the Place Group. A two-way ANOVA revealed a main effect for both consonants [$F(7, 350) = 18.63, p < 0.001$] and condition [$F(5, 350) = 309.22, p < 0.001$], as well as a significant consonant X condition interaction [$F(35, 350) = 6.62, p < 0.001$]. A Newman-Keuls post-hoc test showed that identification scores for the voiceless alveolar fricative [s] were significantly lower than those of all other consonants.

Figure 4 shows clearly that scores for all consonants increased with the duration condition and this increase was dramatic between the 10% and 20% conditions. Scores for the voiceless fricative [s] were strikingly lower at every condition than those of all other consonants, except [d] at the 40% condition.

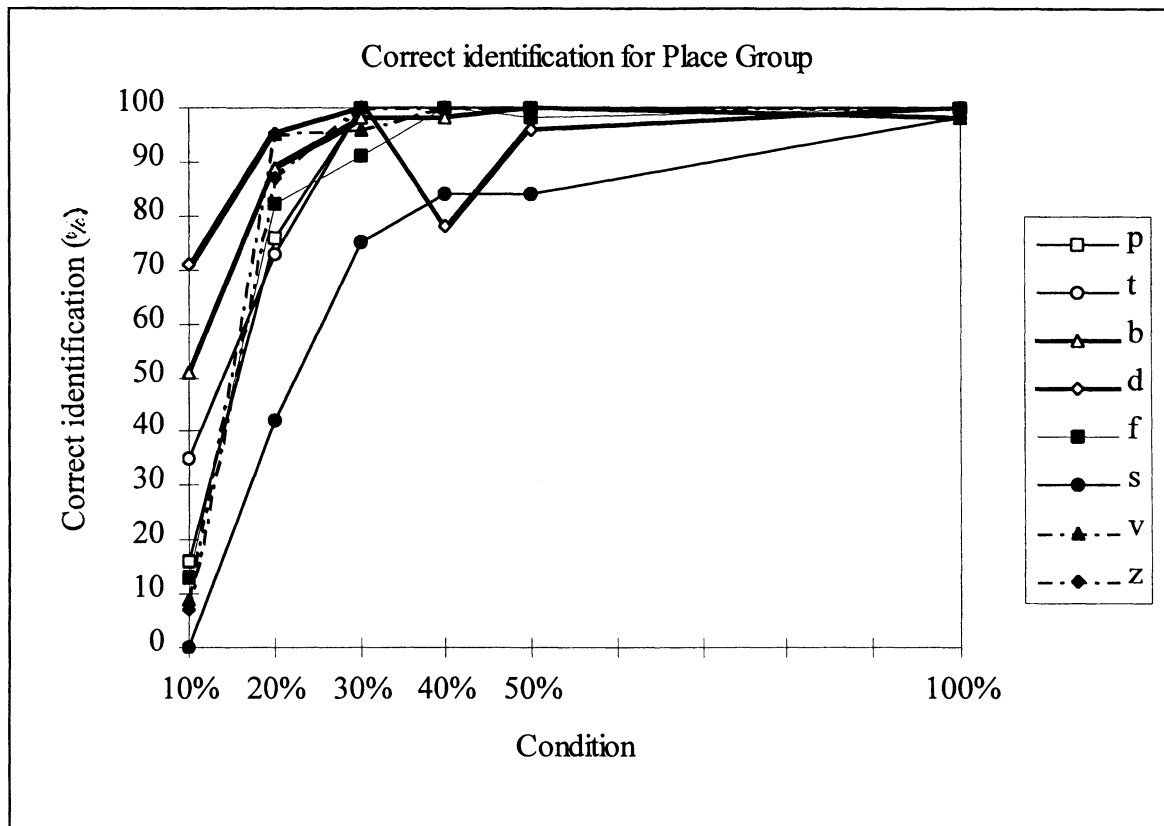


Figure 4. Correct identification for each consonant for each condition for the 11 subjects in the Place Group.

Figure 5 shows the correct identification scores for each consonant across subjects in the Manner Group. A two-way ANOVA revealed a main effect for both consonants [$F(7, 315) = 11.1, p < 0.001$] and condition [$F(5, 325) = 110.89, p < 0.001$]. There was also a significant consonant X condition interaction [$F(35, 315) = 1.99, p < 0.001$]. A Newman-Keuls post-hoc test showed that identification scores for the voiceless alveolar stop [t] were significantly higher than those of the voiceless fricatives, [f] and [s], and the voiced labial stop [b]. Scores for the voiced fricative [v] were also significantly higher than the voiceless fricatives [f] and [s].

Figure 5 shows that the scores for all consonants for the Manner Group increased more gradually with duration condition, than those of the Voicing Group or the Place Group, and there was more variance between consonants at each condition. Scores for the voiceless fricatives, [f] and [s] were strikingly lower at the 30% condition than those of all other consonants.

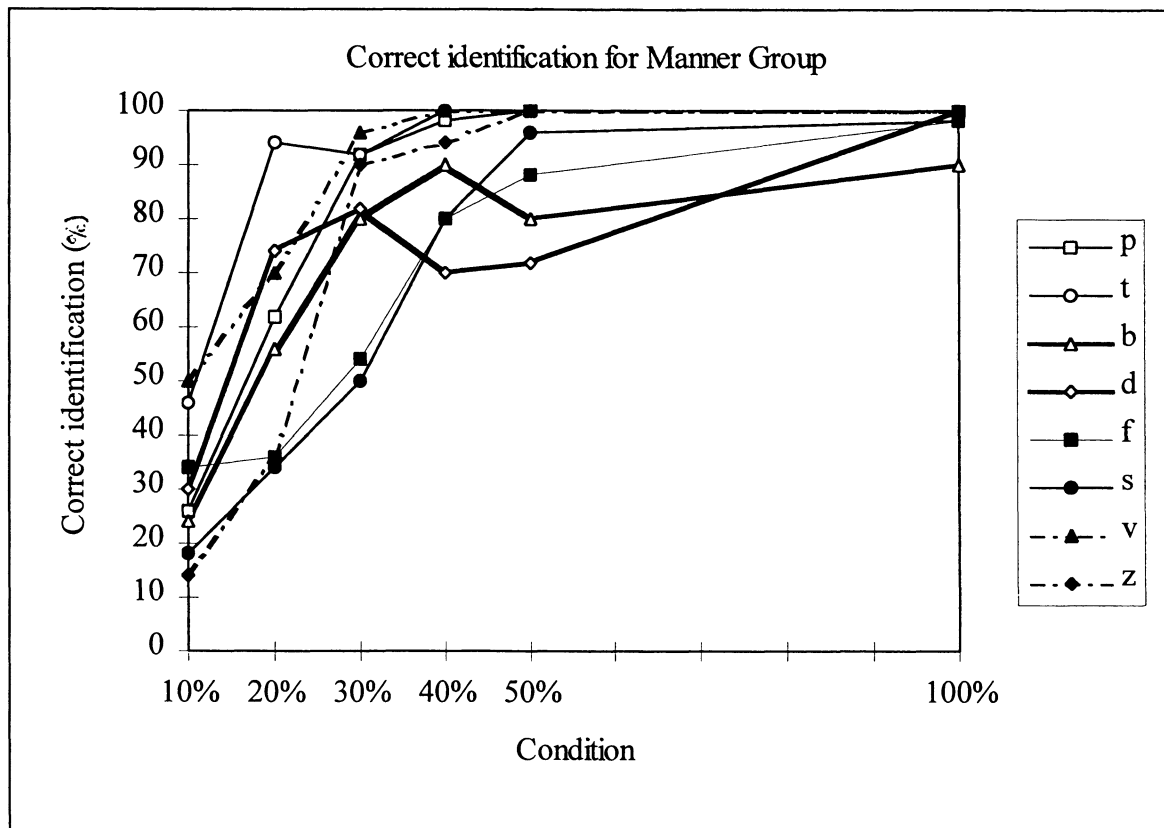


Figure 5. Correct identification for each consonant for each condition for the 10 subjects in the Manner Group.

4.2.2 Identification of voicing feature

Table 7 shows the scores, averaged over the 11 subjects in the Place Group (F_p) and the 10 subjects in the Manner Group (F_m), for correct identification of voicing, for each condition (the Voicing Group is not included since they did not have a voicing decision to make and therefore had 100% correct voicing scores). Two-way ANOVAs were performed on responses from each group to compare the different consonants and conditions. For each group, there was a main effect for both consonant [$F_p(7,480) = 7.94$, $p < 0.001$; $F_m(7,432) = 9.41$, $p < 0.001$] and duration [$F_p(5,480) = 83.9$, $p < 0.001$; $F_m(5,432) = 60.35$, $p < 0.001$]. There was also a significant consonant X condition interaction [$F_p(35,480) = 3.89$, $p < 0.001$; $F_m(35,432) = 2.27$, $p < 0.001$].

A Newman-Keuls post-hoc test revealed that for the Place Group, voicing scores for the voiceless alveolar fricative [s] were significantly lower than scores for any of the voiced consonants [b, d, v, z]. Voicing scores for the voiceless labial fricative [f] were also significantly lower than the scores for [z].

For the Manner Group, a Newman-Keuls post-hoc test showed that voicing scores for both voiceless fricatives, [f] and [s], were significantly lower than scores for the voiced fricatives, [v] and [z], the voiced labial stop [b], and the voiceless alveolar stop [t]. Voicing scores for [t] were also significantly higher than those for its voiced counterpart [d].

Overall, voicing scores for the Place Group were quite high for all consonants by the 20% condition (with a mean score of 85%) whereas voicing scores for the Manner Group rose more gradually.

| Group | Condition | Correct Identification of Voicing (%) | | | | | | | | mean |
|--------|-----------|---------------------------------------|-----|-----|-----|-----|-----|-----|-----|------|
| Place | | [p] | [t] | [b] | [d] | [f] | [s] | [v] | [z] | |
| | 10% | 29 | 38 | 65 | 76 | 27 | 58 | 69 | 80 | 55 |
| | 20% | 76 | 75 | 89 | 95 | 82 | 73 | 95 | 95 | 85 |
| | 30% | 100 | 100 | 98 | 100 | 91 | 75 | 96 | 100 | 95 |
| | 40% | 100 | 100 | 98 | 78 | 100 | 84 | 100 | 100 | 95 |
| | 50% | 100 | 100 | 100 | 96 | 98 | 84 | 100 | 100 | 97 |
| Manner | 100% | 100 | 100 | 98 | 100 | 100 | 98 | 100 | 100 | 100 |
| | 10% | 40 | 64 | 54 | 52 | 40 | 46 | 60 | 62 | 52 |
| | 20% | 62 | 94 | 72 | 80 | 40 | 38 | 76 | 72 | 67 |
| | 30% | 94 | 96 | 94 | 90 | 60 | 56 | 98 | 96 | 86 |
| | 40% | 98 | 100 | 94 | 70 | 88 | 86 | 100 | 96 | 92 |
| | 50% | 100 | 100 | 100 | 74 | 92 | 98 | 100 | 100 | 96 |
| | 100% | 100 | 100 | 100 | 100 | 100 | 98 | 100 | 100 | 100 |

Table 7. Correct identification (in percent) of voicing, for each consonant, for each condition for the Place Group (N = 11) and Manner Group (N = 10).

4.2.3 Identification of place feature

Table 8 shows the scores, averaged over the 10 subjects in the Voicing Group (F_v) and the 10 subjects in the Manner Group (F_m), for correct identification of place, for each condition. Two-way ANOVAs were performed on responses from each group to compare the different consonants and conditions. For each group, there was a main effect for both consonant [$F_v(7,432) = 11.28$, $p < 0.001$; $F_m(7,432) = 6.09$, $p < 0.001$] and duration [$F_v(5,432) = 29.29$, $p < 0.001$; $F_m(5,432) = 50.69$, $p < 0.001$]. There was also a significant consonant X condition interaction [$F_v(35,432) = 3.24$, $p < 0.001$; $F_m(35,432) = 2.58$, $p < 0.001$].

A Newman-Keuls post-hoc test revealed that for the Voicing Group, place scores for the voiced labial stop [b] were significantly lower than scores for any of the other stop consonants [p], [t], or [d], and the voiced labial fricative [v]. Place scores for the fricatives [f] and [z] were also significantly lower than the scores for [p] and [v].

For the Manner Group, a Newman-Keuls post-hoc test showed that place scores for the voiced fricative [z] were significantly lower than scores for the voiceless stop, [p] and [t], and for the voiced fricative [v].

In general, place scores for both groups of subjects were quite high for all consonants except [z] by the 20% condition (mean scores of 83% and 79% for the Voicing Group and the Manner Group respectively).

| Group | Condition | Consonant | | | | | | | | mean |
|---------|-----------|-----------|-----|-----|-----|-----|-----|-----|-----|------|
| Voicing | | [p] | [t] | [b] | [d] | [f] | [s] | [v] | [z] | |
| | 10% | 68 | 76 | 56 | 50 | 40 | 86 | 98 | 26 | 63 |
| | 20% | 100 | 86 | 66 | 84 | 84 | 86 | 100 | 60 | 83 |
| | 30% | 100 | 100 | 44 | 100 | 84 | 86 | 100 | 100 | 89 |
| | 40% | 100 | 100 | 76 | 100 | 84 | 84 | 100 | 100 | 93 |
| | 50% | 100 | 100 | 78 | 100 | 96 | 90 | 100 | 100 | 96 |
| Manne | 100% | 100 | 100 | 100 | 100 | 96 | 100 | 100 | 100 | 100 |
| | 10% | 58 | 70 | 52 | 50 | 82 | 42 | 78 | 24 | 57 |
| | 20% | 86 | 100 | 84 | 82 | 76 | 66 | 94 | 44 | 79 |
| | 30% | 98 | 94 | 86 | 86 | 84 | 88 | 96 | 84 | 90 |
| | 40% | 100 | 100 | 96 | 98 | 92 | 94 | 100 | 98 | 97 |
| | 50% | 100 | 100 | 78 | 100 | 96 | 90 | 100 | 100 | 96 |
| | 100% | 100 | 100 | 100 | 100 | 96 | 100 | 100 | 100 | 100 |

Table 8. Correct identification (in percent) of place, for each consonant, for each condition for the Voicing Group (N = 10) and the Manner Group (N = 10).

4.2.4 Identification of manner feature

Table 9 shows the scores, averaged over the 10 subjects in the Voicing Group (F_v) and the 11 subjects in the Place Group (F_p), for correct identification of manner, for each condition. Two-way ANOVAs were performed on responses from each group to compare the different consonants and conditions. For each group, there was a main effect for both consonant [$F_v(7,432) = 18.37$, $p < 0.001$; $F_p(7,480) = 31.8$, $p < 0.001$] and duration [$F_v(5,432) = 105.94$, $p < 0.001$; $F_p(5,480) = 331.23$, $p < 0.001$]. There was also a

significant consonant X condition interaction [$F_v(35,432) = 6.76, p < 0.001$; $F_p(35,480) = 23.1, p < 0.001$].

A Newman-Keuls post-hoc test revealed that for the Voicing Group, manner scores for the voiceless alveolar fricative [s] were significantly lower than scores for all other consonants. In addition, manner scores for the voiceless labial fricative [f] were also significantly lower than scores for the voiceless stops, [p] and [t], and for the voiced stop [d].

For the subjects in the Place Group, a Newman-Keuls post-hoc test showed that manner scores for the voiceless alveolar fricative [s] were significantly lower than scores for all stop consonants, [p, t, b, d]. Also, manner scores for the voiced fricatives, [v] and [z], were significantly lower than scores for the alveolar stops, [t] and [d].

In general, manner scores for both groups of subjects were quite high for all consonants except [s] by the 20% condition (mean scores of 82% and 95% for the voicing group and the place group respectively). For both groups, mean scores across consonants are at chance at the 10% condition, although it should be noted that [t] was correctly identified as a stop with near perfect accuracy even at the 10% condition.

| Group | Condition | Consonant | | | | | | | | mean |
|---------|-----------|-----------|-----|-----|-----|-----|-----|-----|-----|------|
| Voicing | | [p] | [t] | [b] | [d] | [f] | [s] | [v] | [z] | |
| | 10% | 78 | 100 | 68 | 76 | 18 | 4 | 16 | 24 | 48 |
| | 20% | 96 | 94 | 92 | 94 | 80 | 28 | 84 | 88 | 82 |
| | 30% | 100 | 100 | 100 | 100 | 90 | 74 | 100 | 100 | 96 |
| | 40% | 100 | 100 | 100 | 100 | 92 | 98 | 100 | 100 | 99 |
| | 50% | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Place | 100% | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 10% | 75 | 93 | 69 | 89 | 36 | 0 | 13 | 13 | 49 |
| | 20% | 100 | 96 | 100 | 98 | 100 | 67 | 100 | 100 | 95 |
| | 30% | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 40% | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 50% | 100 | 100 | 100 | 100 | 98 | 100 | 100 | 100 | 100 |
| | 100% | 100 | 100 | 98 | 100 | 100 | 100 | 100 | 100 | 100 |

Table 9. Correct identification (in percent) of manner, for each consonant, for each condition for the Voicing Group (N = 10) and the Place Group (N = 11).

4.2.5 Comparative feature analysis

One aim of this study was to determine whether giving subjects information about a particular feature facilitated their perception of a second feature relative to a third. In order to answer this question, an ANOVA was performed on the results for each group to compare the previous findings of correct identification of each of the features. That is, for each group of subjects, the data in Tables 7, 8, and 9 were subjected to an ANOVA to discover if there was any effect for feature identification. No main effect for feature was found for the Voicing Group or the Place Group. That is, there is no significant difference between the identification of place or manner for the Voicing Group, and no significant difference between the identification in terms of voicing or manner for the Place Group. For the Manner Group, however, a main effect for feature was found [$F(1, 958) = 6.56$, $p < 0.05$]. Scores for identification of voicing were found to be significantly lower than those of identification of place.

4.2.6 Information transmission analysis

The data from each group of subjects were subjected to information transmission analysis to examine any possible response bias. Of interest was the comparison between relative information transmitted for each of the linguistic features: voicing, place, and manner. Figures 6 through 8 show the relative information transmitted for the relevant features for each condition. In general, the percentage of information transmitted for all three features increased as duration increases, with the largest increase for all features occurring between the 20% and 30% condition.

Figure 6 shows that for the Voicing Group, there was more information transmitted about manner than about place at the 30%, 40%, and 50% conditions. These results were not consistent with the results of the previous analyses in that subjects in the Voicing Group showed no significant difference in their ability to perceive manner information relative to place information.

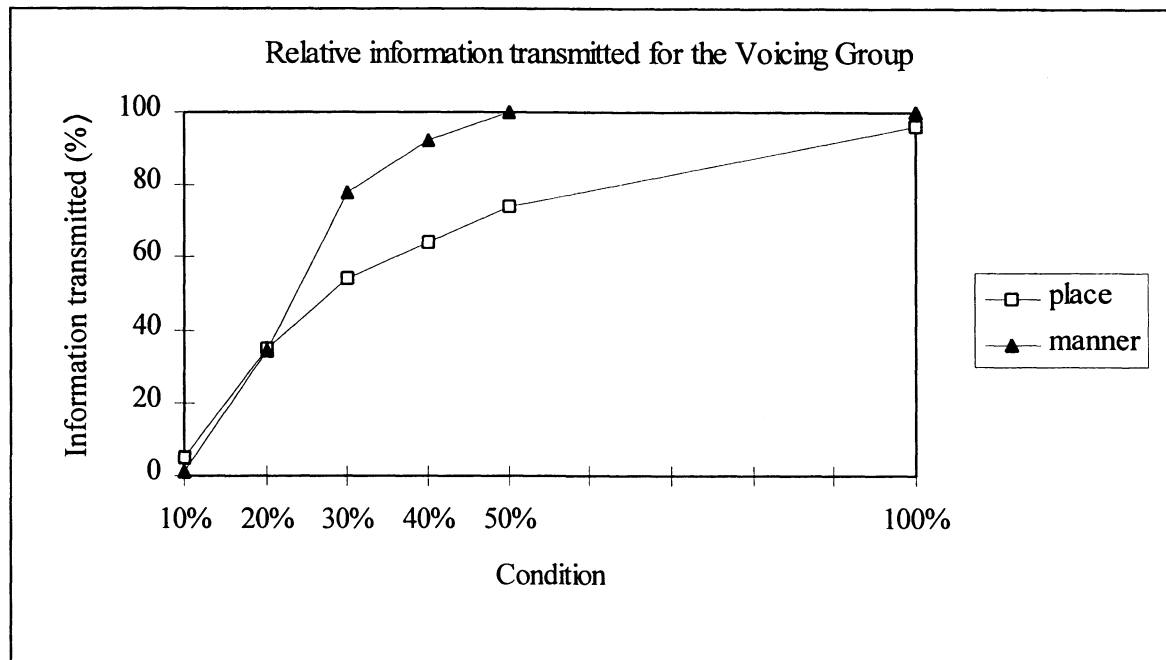


Figure 6. Relative information transmitted (in percent) for place and manner for the Voicing Group.

Figure 7 shows that for the Place Group, there was more information transmitted about manner than voicing. This result was consistent with the previously reported mean scores for identification of voicing and manner in that the scores for manner were higher than the scores for voicing. This difference, however, was not found to be statistically significant.

Figure 8 shows that for the Manner Group, there was more information transmitted about place than about voicing for the first four conditions. This was consistent with the previous findings showing that subjects in the Manner Group were significantly better at identification of place than voicing.

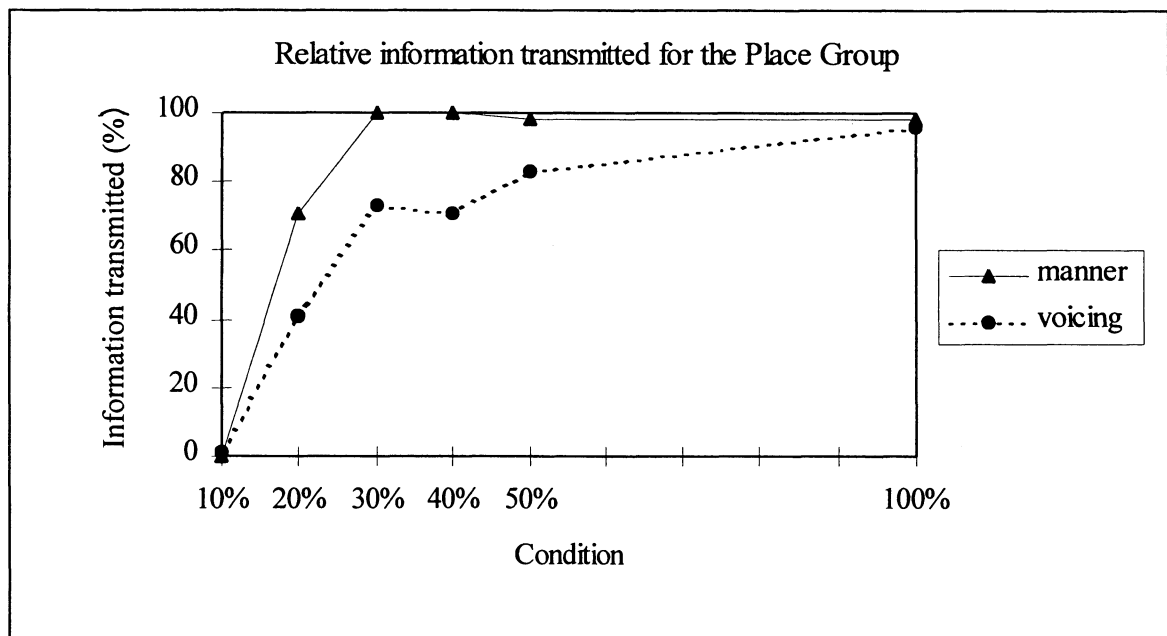


Figure 7. Relative information transmitted (in percent) for voicing and manner for the Place Group.

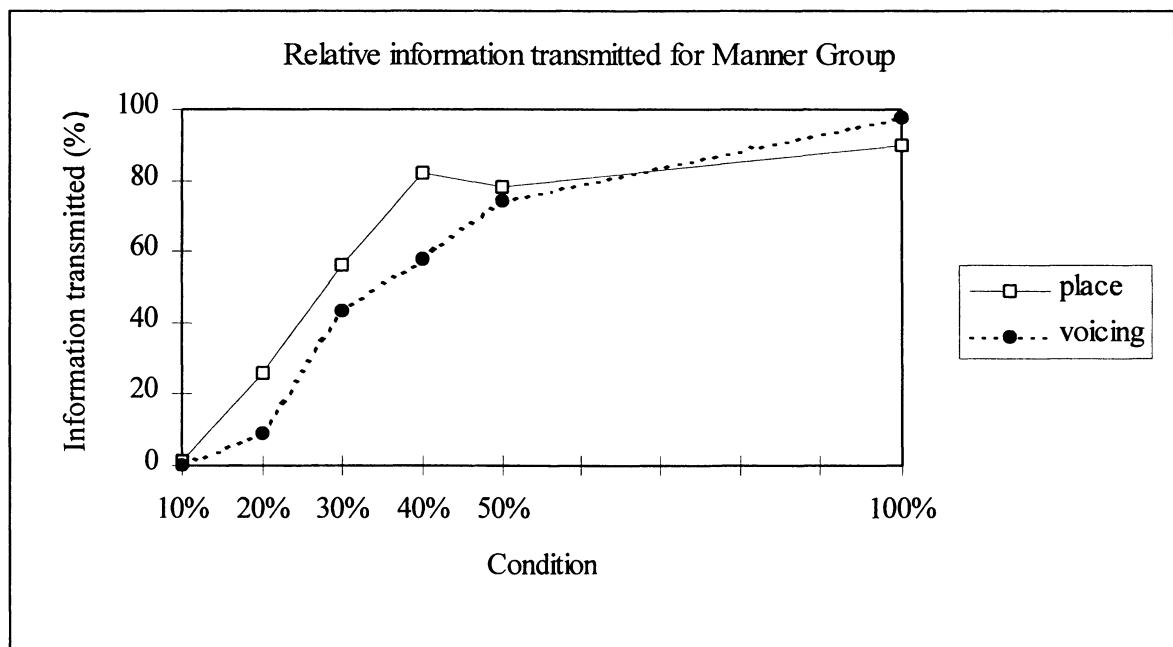


Figure 8. Relative information transmitted (in percent) for voicing and place for the Manner Group.

4.3 Discussion

Experiment 2 was designed to further investigate the perception of features in the identification of English consonants. One finding, which is consistent with the results of Experiment 1, was that correct identification scores for these eight consonants varied significantly. As in Experiment 1, identification scores for the voiceless alveolar stop [t] were significantly higher than scores for the voiceless fricatives [f] and [s], whereas scores for [s] were the lowest. Additionally, in Experiment 2, scores for [t] were also significantly higher than scores for [z] and [b], and scores for [b] and [s] were significantly lower than scores for any other consonant. The comparative feature analysis showed that for the Manner Group, identification of voicing was significantly worse than identification of place. This finding is supported by the information transmission analysis for the Manner Group: more information about place than voicing was transmitted during the shortest four conditions. For the Voicing Group and the Place Group, more information was transmitted about manner than place or voicing, respectively, but this finding did not obtain in the feature identification results. That is, there were no significant differences found between the identification of manner and the voicing or place for these two groups.

5 Overall analysis

In order to examine the effect of providing subjects with certain feature information, the results of Experiments 1 and 2 were analyzed comparatively, following the same basic organization of the previous results sections. For these comparisons, subjects in Experiment 1 will be referred to as the Experiment 1 Group. First, correct identification results of all four groups were averaged across consonants and compared. Then, identification of each of the features, voicing, place and manner, was compared. Where appropriate, ANOVAs were carried out and results are reported. Finally, information transmission analyses of all data is presented for comparison.

5.1 Correct identification analysis

Figure 9 shows the mean correct identification scores (in percent) across consonants for the different conditions for the different subject groups. For the 10% condition, mean correct identification was at chance levels for all groups. (For the Experiment 1 Group, chance was approximately 12%, or one in eight; for the other groups, chance was 25% correct, or one in four.) But for all other conditions, the mean correct responses for the different subject groups were all above chance, and, more importantly, differed from one another. If identification scores had improved in a similar manner for all three groups in

Experiment 2, then this improvement might have simply been due to the overall reduction in the number of stimuli presented at a time and the number of possible responses. That this was not the case is illustrated by a one-way ANOVA, which revealed a main effect for group [$F(3, 1439) = 24.05, p < 0.001$], and a Newman-Keuls post-hoc test which showed that the identification scores of the Place Group were significantly higher than the scores of both the Manner Group and the Experiment 1 Group. Thus the Place Group performed significantly better at consonant identification than the Manner Group or the Experiment 1 Group, despite the fact that all groups identified exactly the same acoustic stimuli. The scores of the Voicing Group were also found to be significantly higher than those of the Experiment 1 Group. Identification scores of the Manner Group, however, were not found to be significantly different from the scores of the Experiment 1 Group. Not having to make a decision about the manner feature, therefore, did not facilitate identification of the consonants.

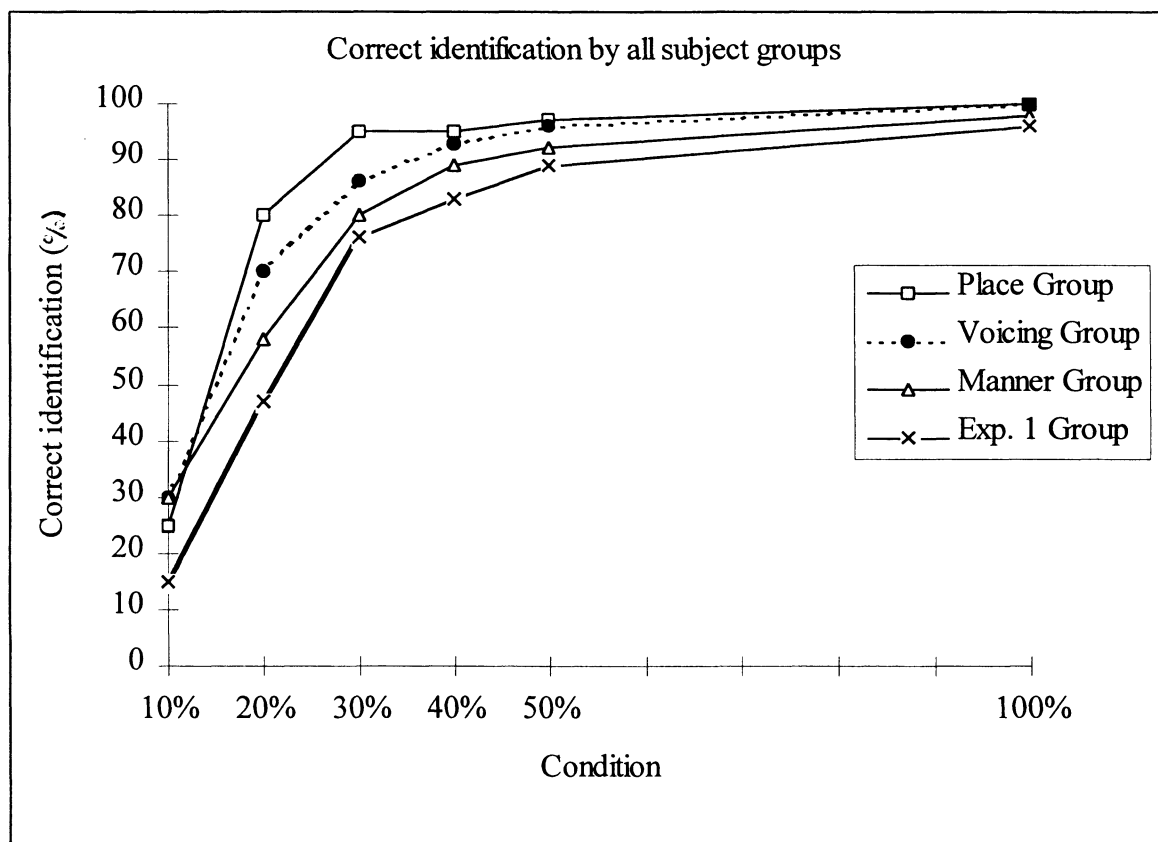


Figure 9. A comparison of correct identification by all subject groups for all conditions.

5.2 Identification of features by all groups

Figure 10 compares the mean scores for correct identification of voicing for the Place Group, the Manner Group, and the Experiment 1 Group. A one-way ANOVA showed a main effect for group [$F(2, 1677) = 6.55, p < 0.01$] and a Newman-Keuls post-hoc test revealed that the Place Group identified voicing significantly better than the Manner Group.

Figure 11 shows the mean scores for correct identification of place for the Voicing Group, the Manner Group, and the Experiment 1 Group. A one-way ANOVA revealed no significant differences among the groups. All groups, therefore, were equally good at identifying the place feature of the consonants, with scores of approximately 80% correct at the 20% condition.

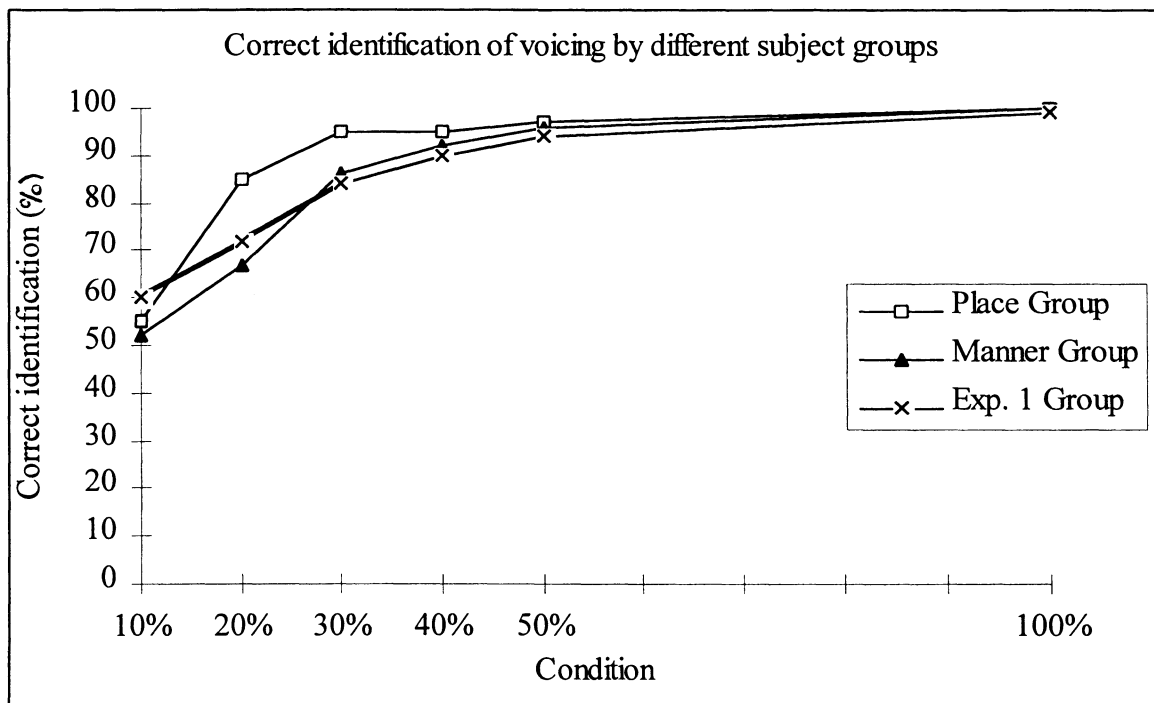


Figure 10. A comparison of correct identification of voicing by different subject groups.

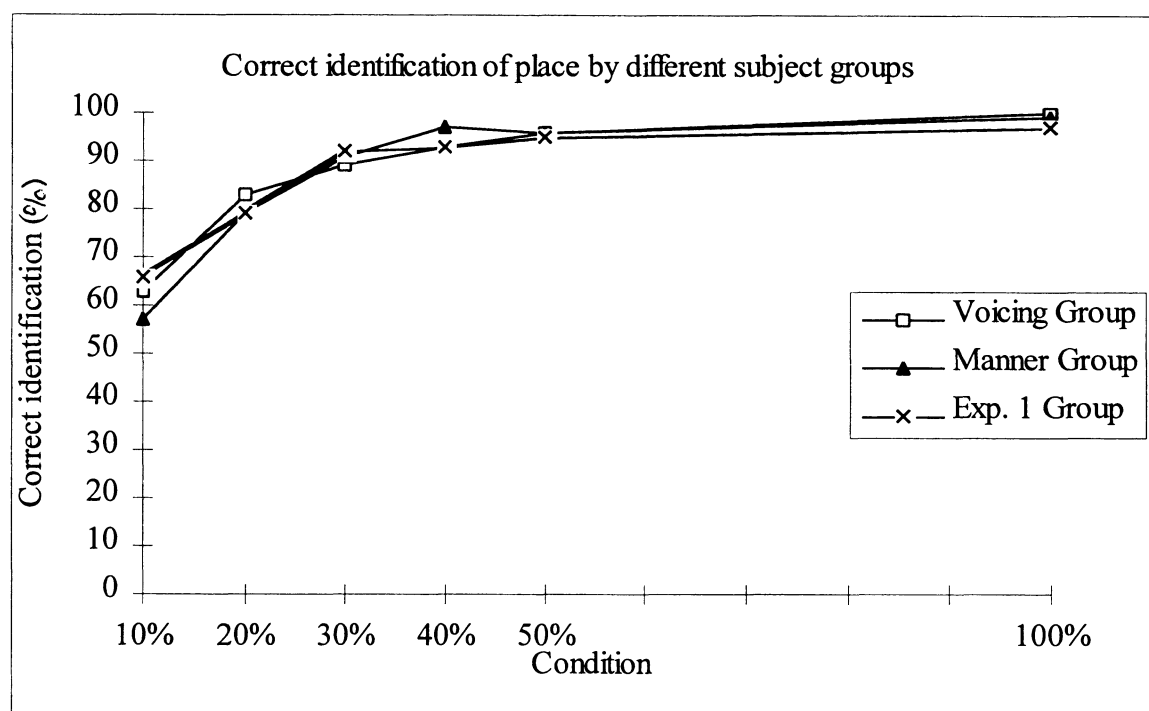


Figure 11. A comparison of correct identification of place by different subject groups.

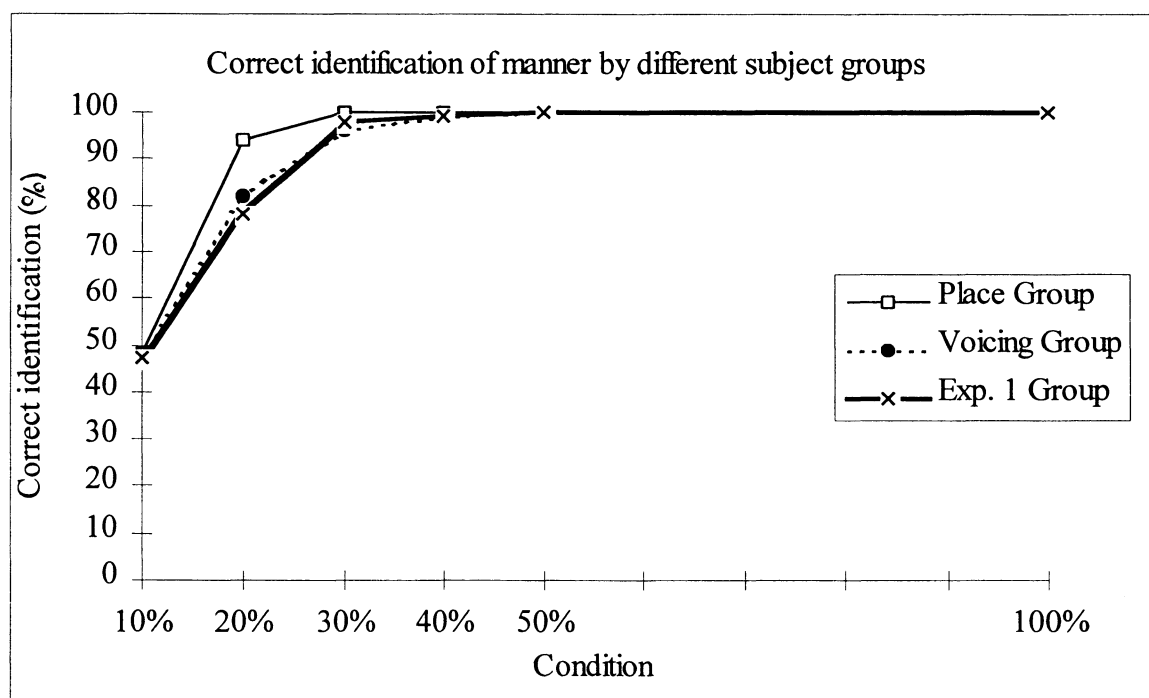


Figure 12. A comparison of correct identification of manner by different subject groups.

Figure 12 shows the mean scores for correct identification of manner across consonants for the Place Group, the Voicing Group, and the Experiment 1 Group. A one-way ANOVA revealed no significant differences among the groups, despite higher scores in the 20% condition for the Place Group.

5.3 Information transmission analysis

Figures 13 through 15 compare the relative information transmitted for voicing, place, and manner, respectively, for the different subject groups. It is clear from these figures that for voicing and manner features, there is more information transmitted for the Place Group than for any other group of subjects. The information transmission analyses of voicing, place, and manner features are similar for the other three groups. The only exception to this generalization is found in Figure 14, where the Manner Group shows more information transmitted about place at the 40% condition than the Voicing Group or Experiment 1 Group.

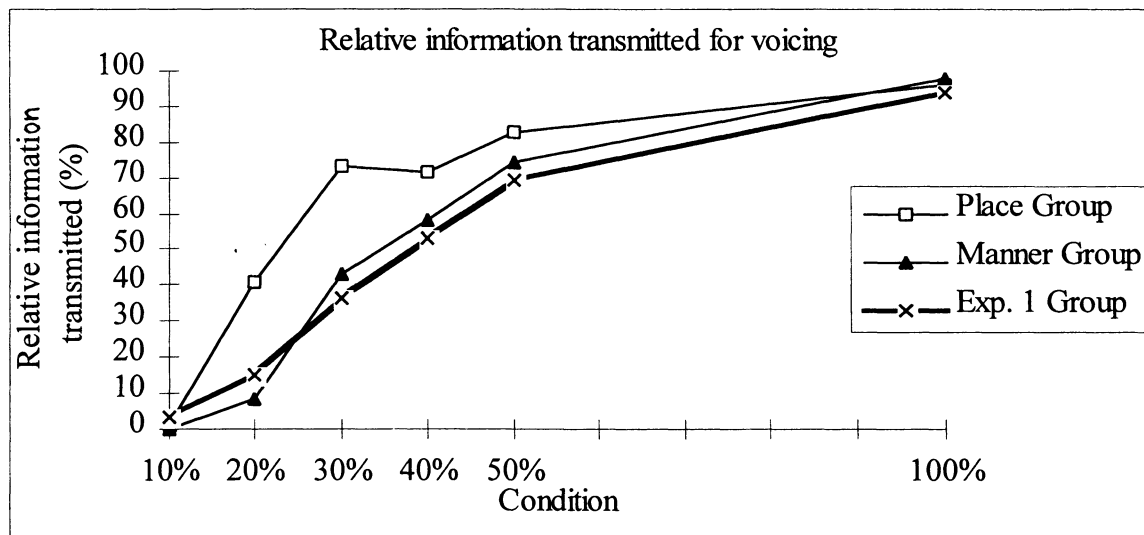


Figure 13. A comparison of relative information transmitted for voicing, for different subject groups.

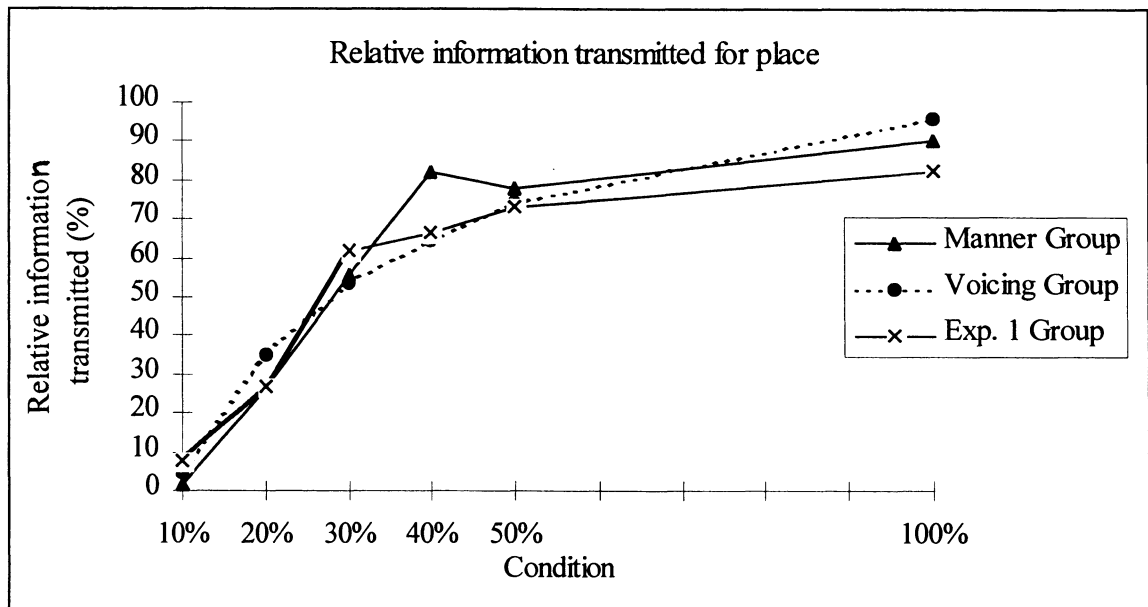


Figure 14. A comparison of relative information transmitted for place, for different subject groups.

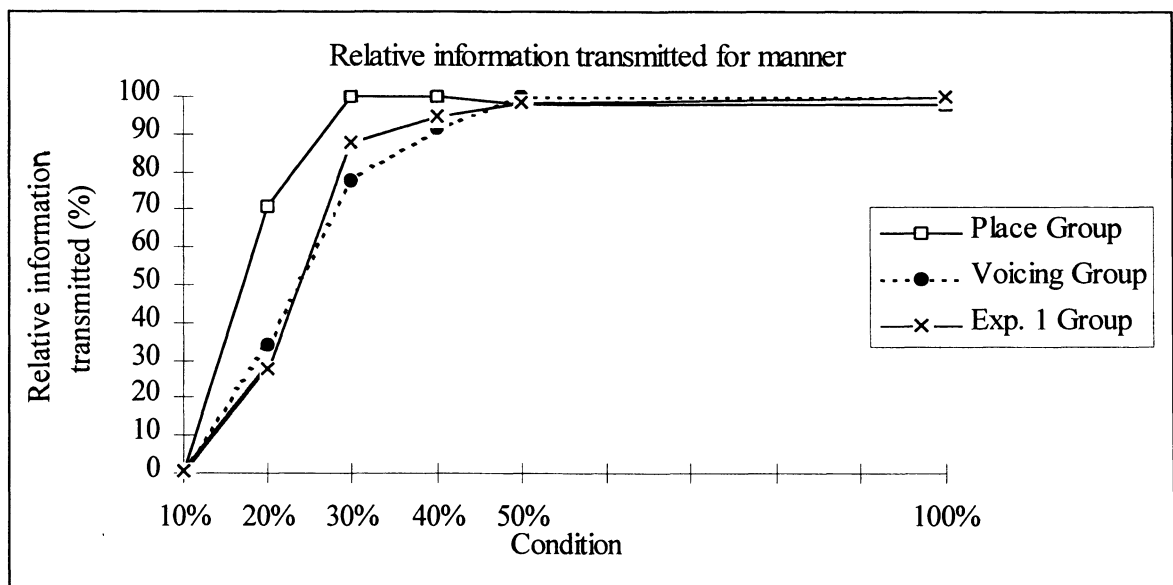


Figure 15. A comparison of relative information transmitted for manner, for different subject groups.

5.4 General discussion

A comparison of the results from Experiment 1 and Experiment 2 supports the hypothesis of an interdependence of feature processing. All four groups of subjects heard the same stimuli, but their performance on an identification task for these stimuli varied significantly, depending upon the composition of the stimulus subsets. The key manipulations in these experiments included the number of stimuli presented, the restricted response choices, and the composition of the stimulus subsets. Evidence that these first two manipulations did not contribute to an increase in identification scores was illustrated by the absence of significant differences between the scores of Manner Group and those of the Experiment 1 Group. Furthermore, significant differences between the scores of the Place Group and the Manner Group can only be due to the composition of the stimulus subsets since both groups heard stimulus sets of the same size. Prior knowledge on the part of the subjects of a particular feature, by way of the composition of the response set, resulted in differences in identification scores.

6 Conclusion

The aim of the experiments in this study was to investigate the role of features in the identification of English consonants. Consonant confusions were elicited and examined to explore possible differences in the relative perceptual saliency of voicing, place, and manner features. In addition, the featural organization of stimulus sets in Experiment 2 was designed to explore possible feature interactions in consonant identification. The hypothesis tested was that differences in identification scores would arise for subjects given different subsets of the same stimuli. The results showed that correct identification of these consonants did indeed depend upon the featural composition of the stimulus subsets. That is, which feature information subjects received and, consequently, which featural decisions had to be made in identification influenced the correct identification of the consonants. Specifically, subjects in the Place Group in Experiment 2 were significantly better at both identification and determination of voicing than the Experiment 1 Group or the Manner Group. This outcome suggests that for the identification of the eight English consonants [p, t, b, d, f, s, v, z], place had a privileged status among the features included in the present study. Place information facilitated identification and reduced ambiguities in voicing and manner features.

It is not the case, however, that subjects performed better when given place information because the place feature was inherently difficult to perceive. This was evident from two sources. First, the feature identification analysis of the data showed that both the Experiment 1 Group and the Manner Group were significantly worse at

perceiving voicing than place. If identification scores improved due to providing subjects with information about a feature that was relatively difficult to perceive, then scores should have improved most for the Voicing Group. This was not the case. Second, there is no evidence of any inherent difficulty in the identification of the place feature, since scores were relatively high for all groups at all conditions other than the 10% condition.

The results of the present study also suggest differences in the perceptual saliency among the features under investigation, but these differences are not entirely consistent with previous findings and may be due in part to the experimental design. It was hypothesized that because voicing seemed to carry more perceptual weight than place features in judgments of similarity, voicing may be more salient to the perceptual system. The greater number of voicing errors compared to place errors in the present study does not support this hypothesis. It may be, however, that in the present study, voicing information was affected more than place or manner information by the gating procedure used, and this was reflected in the correct identification scores of each of the features investigated.

The fact that more voicing confusions than place or manner confusions were found for the Experiment 1 Group also differs from the findings of Miller and Nicely (1955). This apparent inconsistency can be explained by taking into account the manner in which the stimuli were presented in the Miller and Nicely study, namely embedded in noise and low-pass filtered. These manipulations most directly affect the acoustic cues correlated with place. The more important finding of the Miller and Nicely study, and one which is upheld by the present study, is the relative independence of the perception of the different features.

The present findings of high identification scores for place and low scores for voicing are consistent with the findings of Smits (1998). In a consonant recognition study using forward and backward gated British English VCV stimuli, Smits found that at the point of consonantal release for stop consonants, more information was transmitted about place than about manner or voicing. A direct comparison of fricatives is more difficult, since Smits' stimuli were intervocalic, but nevertheless, less information about voicing was transmitted for fricatives throughout the consonant duration.

Smits proposes that in consonant recognition, manner is classified first, and that a determination of where to extract information about place and voicing depends on the result of the manner classification. This implies an interaction of features such that classification of place and voicing is dependent to some extent on manner information. This conclusion does not seem to be supported by the present study, in which the Manner Group performed no better at the identification task than the Experiment 1 Group. If, in

the time course of consonant recognition, processing proceeds from a manner determination to a voicing and place determination, then the identification scores for the subjects in the Manner Group, for whom the manner feature was evident from the response choices, should have been higher than the other groups. It may be, however, that the process of consonant recognition differs for initial versus intervocalic stimuli.

Further research is needed to examine more closely the nature of the feature interactions found in the present study. One possible explanation for the higher identification scores for the Place Group is that these scores were a result of phonetic priming of place information. This hypothesis might account for the data presented here, and yet it still leaves open the question of why place information should have a privileged status in the identification of these phonemes relative to voice or manner information. A more complete picture is needed of the role of distinctive features in both the nature of the internal representation of these phonemes as well as the processes involved in accessing those representations.

7 References

- Blumstein, S. E. (1974) The use and theoretical implications of the dichotic technique for investigating distinctive features. *Brain and Language* 1, 337-350.
- Blumstein, S. E., and Stevens, K. N. (1980) Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America* 67, 648-662.
- Carden, G., Levitt, A., Jusczyk P. W. and Wally, A. (1981) Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception and Psychophysics* 29, 26-36.
- Cole, R. A. (1973) Listening for mispronunciations: A measure of what we hear during speech. *Perception and Psychophysics* 13, 153-156.
- Crandell, C. (1992) Individual speech-recognition susceptibility to noise in elderly listeners. *Journal of the Acoustical Society of America* 92, 2339, 2pSP5.
- Dubno, J. R. and Levitt, H. (1981) Predicting consonant confusions from acoustic analysis. *Journal of the Acoustical Society of America* 69, 249-261.
- Gelfand, S. A., Piper N., and Silman, S. (1985) Consonant recognition in quiet as a function of aging among normal hearing subjects. *Journal of the Acoustical Society of America* 78, 1198-1205.
- Greenberg, J. H. and Jenkins, J. J. (1964) Studies in the psychological correlates of the sound system of American English. *Word* 20, 157-177.

- Hayden, M. E., Kirstein, E., and Singh, S. (1979) Role of distinctive features in dichotic perception of 21 English consonants. *Journal of the Acoustical Society of America* 65, 1039-1046.
- Hertz, S. R. (1991) Streams, phones and transitions: toward a new phonological and phonetic model of formant timing. *Journal of Phonetics* 19, 91-109.
- Jongman, A. (1989) Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America* 85, 1718-1725.
- Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America* 61, 1337-1351.
- Marslen-Wilson, W. D. and Welsh, A. (1978) Processing interaction and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10, 29-63.
- Mertus, J. (1989) *BLISS Manual*. Providence, RI: Brown University.
- Miller, G. A. and Nicely, P. E. (1955) An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27, 623-638.
- Miller, J.L. (1977) Nonindependence of feature processing in initial consonants. *Journal of Speech and Hearing Research* 20, 519-528.
- Mohr, B. and Wong W. S.-Y. (1968) Perceptual distance and the specification of phonological features. *Phonetica* 18, 31-45.
- Peters, R. W. (1963) Dimensions of perception for consonants. *Journal of the Acoustical Society of America* 35, 1985-1989.
- Pisoni, D. B. and McNabb, S. D. (1974) Dichotic interactions of speech sounds and phonetic feature processing. *Brain and Language* 1, 351-362.
- Pols, L. C. W., and Schouten, M. E. H. (1978) Identification of deleted consonants. *Journal of the Acoustical Society of America* 64, 1333-1337.
- Smits, R. (submitted) Temporal distribution of information for human consonant recognition in VCV utterances.
- Studdert-Kennedy, M. and Shankweiler, D. (1970) Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America* 48, 579-594.
- Wish M., and Carroll J. D. (1974) Application of individual differences scaling to studies of human perception and judgment. In E. Carterette and M. Friedman (eds.) *Handbook of Perception* vol. II. NY: Academic Press, 449-491.

8. Appendix A: Confusion Matrices

8.1 Confusion matrices for Experiment 1.

| | p | t | b | d | f | s | v | z | n/a |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|------------|
| p | 6 | | 45 | 9 | 2 | 3 | 3 | 1 | 1 |
| t | 2 | 24 | 4 | 30 | 4 | 2 | 3 | 1 | |
| b | 3 | 2 | 25 | 23 | 7 | 3 | 6 | | 1 |
| d | 3 | 4 | 33 | 20 | 5 | 2 | 2 | 1 | |
| f | 5 | 9 | 19 | 25 | 7 | 1 | 3 | 1 | |
| s | | 40 | 28 | | 1 | 1 | | | |
| v | 2 | 1 | 52 | 8 | 3 | 3 | 1 | | |
| z | 2 | 6 | 36 | 18 | 3 | 3 | 1 | 1 | |

Table 10. Confusion matrix for 10% condition.

| | p | t | b | d | f | s | v | z | n/a |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|------------|
| p | 30 | 2 | 32 | 2 | 1 | | 2 | | 1 |
| t | 3 | 52 | 1 | 9 | 2 | 2 | | 1 | |
| b | 11 | 6 | 35 | 14 | 2 | 1 | 1 | | |
| d | 3 | 8 | 5 | 47 | 1 | 4 | 2 | | |
| f | 3 | 4 | 4 | 7 | 22 | 2 | 26 | 1 | 1 |
| s | 2 | 34 | 3 | | 9 | 15 | 1 | 5 | |
| v | 10 | 2 | 10 | 4 | | | 44 | | |
| z | 2 | 4 | 3 | 9 | 4 | 1 | 27 | 20 | |

Table 11. Confusion matrix for 20% condition.

| | p | t | b | d | f | s | v | z | n/a |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|------------|
| p | 68 | 1 | 1 | | | | | | |
| t | | 70 | | | | | | | |
| b | 12 | | 38 | 19 | | | 1 | | |
| d | | 28 | | 42 | | | | | |
| f | | | | 1 | 34 | 4 | 27 | 4 | |
| s | | 7 | | | 9 | 37 | 1 | 16 | |
| v | | | | 1 | | | 69 | | |
| z | | | | | | 1 | 3 | 66 | |

Table 12. Confusion matrix for 30% condition.

| | p | t | b | d | f | s | v | z | n/a |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|------------|
| p | 70 | | | | | | | | |
| t | | 70 | | | | | | | |
| b | 1 | | 43 | 26 | | | | | |
| d | | 22 | 2 | 46 | | | | | |
| f | | 2 | | | 46 | 2 | 20 | | |
| s | | 1 | | 1 | 5 | 51 | | 12 | |
| v | | | | | | | 70 | | |
| z | | | | | | | | 70 | |

Table 13. Confusion matrix for 40% condition.

| | p | t | b | d | f | s | v | z | n/a |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|------------|
| p | 69 | 1 | | | | | | | |
| t | | 70 | | | | | | | |
| b | 5 | | 48 | 16 | 1 | | | | |
| d | | 21 | | 49 | | | | | |
| f | | | | | 59 | 7 | 4 | | |
| s | | | | | 1 | 67 | | 2 | |
| v | | | | | | | 70 | | |
| z | | | | | | | 3 | 67 | |

Table 14. Confusion matrix for 50% condition.

| | p | t | b | d | f | s | v | z | n/a |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|------------|
| p | 69 | 1 | | | | | | | |
| t | | 70 | | | | | | | |
| b | | | 67 | 3 | | | | | |
| d | | | | 70 | | | | | |
| f | | | | | 56 | 11 | 2 | 1 | |
| s | | | | | | 70 | | | |
| v | | | | | | | 68 | 2 | |
| z | | | | | | 1 | | 69 | |

Table 15. Confusion matrix for 100% condition.

8.2 Confusion matrices for the Voicing Group.

| | p | t | f | s | n/a |
|----------|----------|----------|----------|----------|------------|
| p | 25 | 14 | 9 | 1 | 1 |
| t | 6 | 38 | 6 | | |
| f | 12 | 28 | 8 | 1 | 1 |
| s | 5 | 43 | 2 | | |

| | b | d | v | z | n/a |
|----------|----------|----------|----------|----------|------------|
| b | 17 | 17 | 11 | 5 | |
| d | 16 | 22 | 9 | 3 | |
| v | 41 | 1 | 8 | | |
| z | 25 | 13 | 12 | | |

Table 16. Confusion matrices for 10% condition.

| | p | t | f | s | n/a |
|----------|----------|----------|----------|----------|------------|
| p | 48 | | 2 | | |
| t | 2 | 43 | 5 | | |
| f | 3 | 7 | 39 | 1 | |
| s | | 36 | 7 | 7 | |

| | b | d | v | z | n/a |
|----------|----------|----------|----------|----------|------------|
| b | 29 | 17 | 4 | | |
| d | 5 | 42 | 3 | | |
| v | 8 | | 42 | | |
| z | 4 | 2 | 16 | 28 | |

Table 17. Confusion matrices for 20% condition.

| | p | t | f | s | n/a |
|----------|----------|----------|----------|----------|------------|
| p | 50 | | | | |
| t | | 50 | | | |
| f | 1 | 4 | 41 | 4 | |
| s | | 13 | 7 | 30 | |

| | b | d | v | z | n/a |
|----------|----------|----------|----------|----------|------------|
| b | 22 | 28 | | | |
| d | | 50 | | | |
| v | | | 50 | | |
| z | | | | 50 | |

Table 18. Confusion matrices for 30% condition.

| | p | t | f | s | n/a |
|----------|----------|----------|----------|----------|------------|
| p | 50 | | | | |
| t | | 50 | | | |
| f | 1 | 3 | 41 | 5 | |
| s | | 1 | 8 | 41 | |

| | b | d | v | z | n/a |
|----------|----------|----------|----------|----------|------------|
| b | 38 | 12 | | | |
| d | | 50 | | | |
| v | | | 50 | | |
| z | | | | 50 | |

Table 19. Confusion matrices for 40% condition.

| | p | t | f | s | n/a |
|---|----|----|----|----|-----|
| p | 50 | | | | |
| t | | 50 | | | |
| f | | | 48 | 2 | |
| s | | | 5 | 45 | |

| | b | d | v | z | n/a |
|---|----|----|----|----|-----|
| b | 39 | 11 | | | |
| d | | 50 | | | |
| v | | | 50 | | |
| z | | | | 50 | |

Table 20. Confusion matrices for 50% condition.

| | p | t | f | s | n/a |
|---|----|----|----|----|-----|
| p | 50 | | | | |
| t | | 50 | | | |
| f | | | 48 | 2 | |
| s | | | | 50 | |

| | b | d | v | z | n/a |
|---|----|----|----|----|-----|
| b | 50 | | | | |
| d | | 50 | | | |
| v | | | 50 | | |
| z | | | | 50 | |

Table 21. Confusion matrices for 100% condition.

8.3 Confusion matrices for Place Group.

| | p | b | f | v | n/a |
|---|----|----|---|----|-----|
| p | 9 | 32 | 7 | 7 | |
| b | 10 | 28 | 9 | 8 | |
| f | 8 | 26 | 7 | 13 | 1 |
| v | 15 | 33 | 2 | 5 | |

| | t | d | s | z | n/a |
|---|----|----|---|---|-----|
| t | 19 | 32 | 2 | 2 | |
| d | 10 | 39 | 3 | 3 | |
| s | 32 | 23 | | | |
| z | 8 | 40 | 3 | 4 | |

Table 22. Confusion matrices for 10% condition.

| | p | b | f | v | n/a |
|---|----|----|----|----|-----|
| p | 42 | 13 | | | |
| b | 6 | 49 | | | |
| f | | | 45 | 10 | |
| v | | | 3 | 52 | |

| | t | d | s | z | n/a |
|---|----|----|----|----|-----|
| t | 40 | 13 | 1 | 1 | |
| d | 2 | 52 | 1 | | |
| s | 17 | 1 | 23 | 14 | |
| z | | 4 | 3 | 48 | |

Table 23. Confusion matrices for 20% condition.

| | p | b | f | v | n/a |
|---|----|----|----|----|-----|
| p | 55 | | | | |
| b | 1 | 54 | | | |
| f | | | 50 | 5 | |
| v | | | 2 | 53 | |

| | t | d | s | z | n/a |
|---|----|----|----|----|-----|
| t | 55 | | | | |
| d | | 55 | | | |
| s | | | 41 | 14 | |
| z | | | | 55 | |

Table 24. Confusion matrices for 30% condition.

| | p | b | f | v | n/a |
|---|----|----|----|----|-----|
| p | 55 | | | | |
| b | 1 | 54 | | | |
| f | | | 55 | | |
| v | | | | 55 | |

| | t | d | s | z | n/a |
|---|----|----|----|----|-----|
| t | 55 | | | | |
| d | 12 | 43 | | | |
| s | | | 46 | 9 | |
| z | | | | 55 | |

Table 25. Confusion matrices for 40% condition.

| | p | b | f | v | n/a |
|---|----|----|----|----|-----|
| p | 55 | | | | |
| b | | 55 | | | |
| f | | | 54 | | 1 |
| v | | | | 55 | |

| | t | d | s | z | n/a |
|---|----|----|----|----|-----|
| t | 55 | | | | |
| d | 2 | 53 | | | |
| s | | | 46 | 9 | |
| z | | | | 55 | |

Table 26. Confusion matrices for 50% condition.

| | p | b | f | v | n/a |
|---|----|----|----|----|-----|
| p | 55 | | | | |
| b | | 54 | | | 1 |
| f | | | 55 | | |
| v | | | | 55 | |

| | t | d | s | z | n/a |
|---|----|----|----|----|-----|
| t | 55 | | | | |
| d | | 55 | | | |
| s | | | 54 | 1 | |
| z | | | | 55 | |

Table 27. Confusion matrices for 100% condition.

8.4 Confusion matrices for the Manner Group.

| | p | t | b | d | n/a |
|----------|----------|----------|----------|----------|------------|
| p | 13 | 7 | 16 | 13 | 1 |
| t | 9 | 23 | 4 | 12 | 2 |
| b | 14 | 6 | 12 | 15 | 3 |
| d | 11 | 10 | 11 | 15 | 3 |

| | f | s | v | z | n/a |
|----------|----------|----------|----------|----------|------------|
| f | 17 | 3 | 24 | 5 | 1 |
| s | 14 | 9 | 14 | 12 | 1 |
| v | 14 | 3 | 25 | 5 | 3 |
| z | 12 | 5 | 24 | 7 | 2 |

Table 28. Confusion matrices for 10% condition.

| | p | t | b | d | n/a |
|----------|----------|----------|----------|----------|------------|
| p | 31 | | 12 | 7 | |
| t | | 47 | | 3 | |
| b | 14 | | 28 | 8 | |
| d | 6 | 4 | 3 | 37 | |

| | f | s | v | z | n/a |
|----------|----------|----------|----------|----------|------------|
| f | 18 | 2 | 20 | 10 | |
| s | 2 | 17 | 15 | 16 | |
| v | 12 | | 35 | 3 | |
| z | 10 | 4 | 18 | 18 | |

Table 29. Confusion matrices for 20% condition.

| | p | t | b | d | n/a |
|----------|----------|----------|----------|----------|------------|
| p | 46 | 1 | 3 | | |
| t | 2 | 46 | 1 | 1 | |
| b | 3 | | 40 | 7 | |
| d | 3 | 2 | 4 | 41 | |

| | f | s | v | z | n/a |
|----------|----------|----------|----------|----------|------------|
| f | 27 | 3 | 15 | 5 | |
| s | 3 | 25 | 3 | 19 | |
| v | | | 48 | 1 | 1 |
| z | | 2 | 3 | 45 | |

Table 30. Confusion matrices for 30% condition.

| | p | t | b | d | n/a |
|----------|----------|----------|----------|----------|------------|
| p | 49 | | 1 | | |
| t | | 50 | | | |
| b | 3 | | 45 | 2 | |
| d | 1 | 14 | | 35 | |

| | f | s | v | z | n/a |
|----------|----------|----------|----------|----------|------------|
| f | 40 | 4 | 6 | | |
| s | 3 | 40 | | 7 | |
| v | | | 50 | | |
| z | | 2 | 1 | 47 | |

Table 31. Confusion matrices for 40% condition.

| | p | t | b | d | n/a |
|---|----|----|----|----|-----|
| p | 50 | | | | |
| t | | 50 | | | |
| b | | | 40 | 10 | |
| d | | 12 | 1 | 36 | 1 |

| | f | s | v | z | n/a |
|---|----|----|----|----|-----|
| f | 44 | 2 | 4 | | |
| s | 1 | 48 | | 1 | |
| v | | | 50 | | |
| z | | | | 50 | |

Table 32. Confusion matrices for 50% condition.

| | p | t | b | d | n/a |
|---|----|----|----|----|-----|
| p | 50 | | | | |
| t | | 50 | | | |
| b | | | 45 | 5 | |
| d | | | | 50 | |

| | f | s | v | z | n/a |
|---|----|----|----|----|-----|
| f | 49 | 1 | | | |
| s | | 49 | | 1 | |
| v | | | 50 | | |
| z | | | | 50 | |

Table 33. Confusion matrices for 100% condition.