REPORT III

EXPERIMENTS

IN SPEECH PERCEPTION

Phonetics Research Seminar

1982 - 1983
PERILUS will mainly contain reports on current work in the Phonetics Laboratory at Stockholm University. Copies are available from the Institute of Linguistics, Stockholm University, S - 106 91 Stockholm, Sweden.
<table>
<thead>
<tr>
<th>CONTENTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION..............................................................1</td>
<td></td>
</tr>
<tr>
<td>Björn Lindblom</td>
<td></td>
</tr>
<tr>
<td>2. ELICITATION AND PERCEPTUAL JUDGEMENT OF DISFLUENCY AND STUTTERING..............................................3</td>
<td></td>
</tr>
<tr>
<td>Anne-Marie Almé</td>
<td></td>
</tr>
<tr>
<td>3. INTELLIGIBILITY VS REDUNDANCY -- CONDITIONS OF DEPENDENCY............................................................27</td>
<td></td>
</tr>
<tr>
<td>Sheri Hunnicutt</td>
<td></td>
</tr>
<tr>
<td>4. THE ROLE OF VOWEL CONTEXT ON THE PERCEPTION OF PLACE OF ARTICULATION FOR STOPS................................45</td>
<td></td>
</tr>
<tr>
<td>Diana Krull</td>
<td></td>
</tr>
<tr>
<td>5. VOWEL CATEGORIZATION BY THE BILINGUAL LISTENER...........81</td>
<td></td>
</tr>
<tr>
<td>Richard Schulman</td>
<td></td>
</tr>
<tr>
<td>6. COMPREHENSION OF FOREIGN ACCENTS (A Cryptic investigation) Richard Schulman and Maria Wingstedt..............101</td>
<td></td>
</tr>
<tr>
<td>7. SYNTETISKT TAL SOM HJÄLPMEDEL VID KORREKTION AV DÖVAS TAL.................................................115</td>
<td></td>
</tr>
<tr>
<td>Anne-Marie Öster</td>
<td></td>
</tr>
</tbody>
</table>
INTRODUCTION

This issue of PERILUS contains papers written by students enrolled in our phonetics program.

The studies by Almé, Hunnicutt and Krull began as "lab projects" in a course on Experimental Phonetics that we offer every year to graduate students and third semester students and whose theme tends to alternate between production, acoustic phonetics and perception (cf. previous issues of PERILUS).

Schulman's contribution and his paper with Wingstedt are investigations undertaken as part of a research project initiated by Tore Janson and sponsored by FRN, "Variationer i talljudsuppfattning".

Öster did her experimental work at Talöverföring, KTH, and submitted this version of her paper as part of the fulfilment of the requirements of the "C"- or third semester course in phonetics.

I thank Jim Lubker for collaborating on the supervision of certain aspects of these projects.

Stockholm in November, 1984

Björn Lindblom
ELICITATION AND PERCEPTUAL JUDGMENT OF DISFLUENCY AND STUTTERING

Ann-Marie Almé

Abstract

One purpose of the present investigation was to find a method for eliciting fluency breakdowns and to examine the distribution and types of disfluencies. A second purpose was to examine how naive listeners would rate fluency breakdowns and to what extent they considered stuttering to be exhibited.

Six stutterers performed three different speaking tasks which were experimentally controlled to elicit fluency breakdowns. The subjects formed two subgroups, which were found to be significantly different in the measures of type, frequency, and distribution of disfluency.

In order to observe how disturbing disfluent speech is to the listener, a listener panel of 20 college students rated randomly selected samples on a 5-point scale. The results indicated that the amount of disfluency and the nature of speech disruptions are important cues in judging the fluency of speech. When judging the overall severity of stuttering the judges were found to be in close agreement with the calculated quantitative analysis. They showed, however, more variability in the correct discrimination between stuttering and normal non-fluency.
ELICITATION AND PERCEPTUAL JUDGMENT OF DISFLUENCY AND STUTTERING

Ann-Marie Almé

The assessment of stuttering is an extremely complex problem. A large number of general issues arise. For example; we must decide what to assess and under what conditions to assess it. We must take into account both the overt and the covert reaction of the stutterer. Added complexity arises from the fact that as a communicative event, a listener is involved and the listener's evaluation of the "stutterer's" speech is an important aspect of the whole process. A great deal of research relevant to such general issues is available in the literature, a limited review of which follows.

Formal research (Langova and Šváb, 1975; Dalton and Hardcastle, 1977; Van Riper, 1982), as well as the author's clinical observation, suggests that variations in fluency appear to be related to the situation in which speech is used, due to situational anxiety and cognitive load.

Adult stutterers have been shown to vary in the type and number of overt symptoms they exhibit, as well as in the frequency with which these symptoms disrupt their speech (Van Riper, 1982). Further, the results of several studies suggest that instances of disfluency are not distributed randomly in the speech of stutterers or non-stutterers (Silverman and Williams, 1967; Schiavetti, 1975; Kaasin and Bjerkan, 1982; and Van Riper, 1982).

From the listener's point of view, some types of disfluency are more readily judged as stuttering than are others. Sound/syllable repetitions and sound prolongations are more often classified as stuttering than other forms of disfluency (Boehmler, 1958; Williams and Kent, 1958). Curlee (1981) found
a substantial variability among listener's perception of disfluency vs stuttering, and low agreement on loci of stuttering. According to Miller and Hewgill (1964) a listener will classify a speaker as a stutterer if he has six disfluencies for every hundred words he produces. It has also been observed that disfluencies which break words or delay their utterance are judged as being more severe than those which leave whole words intact (Schiavetti, 1975).

Moving to the problem of deciding under what conditions to assess the stuttering behavior, one problem is the difficulty in obtaining an overall view of the stutterer's "real" speech pattern in the relatively contrived situation of a laboratory experiment. As noted above, a given individual's stuttering behavior or disfluency, is likely to be so variable that it is important not to limit the experimental speech tasks to oral reading only. There are some severe stutterers who can read with remarkable fluency. The stutterer's performance in oral reading may thus not be representative of his real difficulty in communication. On the other hand, oral reading tasks provide a more easily controlled comparison than does spontaneously evoked speech. Thus, decisions regarding assessment conditions and techniques become crucial.

Little research has been published in Sweden on issues such as those discussed above. Therefore, one purpose of the present study was to find a method for eliciting fluency breakdowns in specific speech situations, and to examine the distribution and types of disfluencies in those situations.

A second purpose of the investigation was to examine how naive listeners would rate fluency breakdowns and to what extent they considered stuttering to be exhibited. That is, we will attempt to determine how disturbing different forms and frequencies of speech disruptions are to naive listeners.

In order to approach these two issues it is important to provide an operational definition of stuttering. In this
investigation the various types of disfluency are classified into nine categories. The categories are based on material presented by Johnson (1961) and Wingate (1964).

PWR - Part-Word Repetition: one or more repetitions of an element equal to or less than a syllable (including monosyllabic words).

AP - Audible Prolongation: audible prolongation of a phonetic segment.

SP - Silent Prolongation: tensely articulated stop closures (blocks).

WWR - Whole-Word Repetitions: one or more repetitions of a word (not including monosyllabic words).

PR - Phrase Repetition: repeated utterance of two or more words.

INT - Interjection: insertion of an extra sound before or within a word.

R - Revision: modification of phrase content.

BW - Broken Word: a word not completely pronounced.

HP - Hesitation Pause: an unusual unfilled pause which could not be classified as a silent prolongation.

The first three categories (PWR, AP, SP) were chosen as the definition of stutterings. "Stuttering interruptions to the forward flow of speech are primarily intra-morphemic; normal disfluencies are primarily supra-morphemic" (Van Riper, 1982). "...two "kernel" characteristics of stuttering speech can be discriminated: (a) repetitions (audible or silent) of single unit speech elements; and (b) prolongations (audible or silent)." (Wingate, 1964). According to the above definition the remaining six categories are considered as normal disfluencies.

These definitions can be applied to four psycholinguistic variables: (1) grammatical class, (2) word length, (3) word position in a phrase, and (4) position within a word. "Grammatical class" was defined as content words (e.g. nouns, verbs and adjectives (Dalton and Hardcastle, 1977) and function
words. "Word length" was defined as the number of syllables in the word. "Word position in a phrase" was defined as each word from the first to the last in each phrase. "Position within a word" was defined as three positions: initial sound, first syllable and medial syllable(s).

By varying the complexity of the speech material and by progressively introducing more communicative demands and emotional stress an increase in disfluency should result. The phenomenon of adaptation, that is the tendency to become more fluent in successive readings of the same or similar passages should thus be avoided.

METHOD

Subjects

One female and five male stutterers ranging in age from 27 to 46 years (mean age 38 years) served as subjects. The severity of overt stuttering varied from mild to severe.

Reading material

Condition I (TXT): oral reading task where the reading material was a 160 word passage from a neutral text.

Condition II (DIA): reading a dialogue (Key-Åberg, 1972) of 200 words. The experimenter was in the same room and read part of the dialogue, but the subjects had the longer lines and had to express the more emotional passages. The text in this dialogue reflected a communicative situation of every day life.

Condition III (SPO): this condition should reflect spontaneous speech to a greater extent than the pre-
Procedure

The subjects were told as little as possible about the experiment or its purpose. They were simply informed that they would be given something to read in their customary manner (i.e. they need not try to control stuttering). Subjects were tested individually in a soundproofed booth and their performances on each session were audio taped. Recordings were made on a Teak R-340 tape recorder at a speed of 19 cm/s. The freestanding microphone (Sennheiser Mo 211 U) was placed at a distance of 35-40 cm from the subjects' mouth.

Five phrases containing 4-9 words were randomly selected from each condition for analysis. The phrases from condition I and II were the same for each subject, but in the third condition they differed to some extent due to the spontaneous nature of that condition. In order to have an exact delimitation of a phrase and the same sound level for all phrases, a special computer program was used (written by R. Carlsson at KTH, Stockholm). The final test tape consisted of 90 phrases with a silent interval of 5 seconds duration between the phrases for a given subject and 10 seconds between subjects. The phrases were randomized for each subject.

Scoring of the disfluencies was then carried out by the experimenter (a certified speech pathologist and phonetician). A count which was identical on three consecutive examinations of the speech samples was accepted as the final score.

For the second part of the experiment 20 college students with normal hearing served as judges. They were seated in a sound treated room and were headphones. The experimental tape consisting of a group of practice samples and the 90 experimental samples was played at a comfortable sound level. The judges recorded their ratings of the speech samples on a res-
response sheet which had a rating scale from 1 - 5. The endpoints were defined as "very disfluent speech" (1) and "very fluent speech" (5). No other points were defined. In the instructions which were given to the judges they were informed that they would hear speech varying in fluency and they were to assign a number between 1 and 5 depending on the degree of fluency. In order to acquaint the judges with the kind of speech samples and the range of fluency/disfluency they would hear, 15 phrases (representing the entire continuum) were used as a training session.

After this first experimental session the judges were asked to listen to the tape again. Now they were instructed to judge the phrases with regard to stuttering. No definition of stuttering was given, they were simply asked to mark on a protocol if they could hear instances of stuttering or not.

RESULTS

Distribution of disfluency in different speech conditions

Counts of each subject's disfluencies were made from the audio tape. The syllable was chosen as the reference unit, in preference to the word, since it is relatively invariant in length, thereby giving a more stable basis for comparison. Frequency counts of disfluency and stuttering are expressed in percentages in relation to the total number of syllables (% SyD and % SyS).

As a group, the subjects showed an increase in disfluency as the speech task become more "difficult" which was in accordance with the expected outcome. The mean frequency of % SyD was 9% (TXT), 12% (DIA), and 19% (SPO). The mean frequency of % SyS was 7% (TXT), 10% (DIA), and 11% (SPO).
Figure 1a and 1b. Mean percentage of disfluency (%SyD) and stuttering (%SyS) for mild stutterers (M) and moderate-severe stuttering (MS) for oral reading, dialogue, and spontaneous speech during three successive reading tasks: I = oral reading, II = dialogue, and III = spontaneous speech.

% SYD
0
10
20
30
40
50
60
70
80
90
100
M-Group
M-Group
MS-Group
MS-Group

% SYS
0
10
20
30
40
50
60
70
80
90
100
M-Group
M-Group
MS-Group
MS-Group
Looking at individual data however, a somewhat different pattern was indicated. The six subjects could be said to form two subgroups. Depending on the frequency of disfluency, three of the subjects were classified as mild stutterers (M-group with a mean of 6.8 % SyD) and three were classified as moderate-severe stutterers (MS-group with a mean of 20.2 % SyD). The main finding concerning the differences in disfluency and stuttering under different conditions is illustrated in Figure la and lb.

The M-group was significantly more disfluent in the third condition ($p < .001$, two-tailed t-test). Of particular interest are data from the MS-group. This group showed a significant increase in disfluency and stuttering from oral reading to dialogue condition ($p < .05$ two-tailed). In condition III they remained highly disfluent but the decrease in stuttering is remarkable. For all six subjects it is evident that the third condition was associated with a substantial increase in % SyD compared to the first condition.

There was a significant between-group difference in condition I and II ($p < .001$). The mean frequency of %SyD and %SyS was higher for the MS-group in the third condition but the difference was not significant.

Type of disfluency

The number of disfluences in each category for both groups are presented in Table I. For the total group 68 % of the disfluencies were classified as stutterings. PWR was found to be the most common speech disruption. The more severe stutterers exhibited substantially more repetition of phonetic segments. They also had instances of stuttering within a word most of them occurring on stressed syllables. Silent and audible prolongations of sounds were found to be the second most common type of disfluency. These types were rarely observed in the speech of mild stutterers (32 % vs 2 %).

16 % of observed disfluencies were interjections. Although insertions of extra sounds or syllables are not uncom-
mon in normal speech, such disruptions are probably sometimes part of the stuttering behavior as they occur as "starters". Starters can be used before a difficult sound or word to facilitate the utterance, to get started. Gradually the starters will be stereotyped and of no aid anymore. One subject had insertion of "uh" sounds before and within several words, but with the stringent operational definition of stuttering in this study they were classified as disfluencies.

Table I. Distribution of type of disfluency for mild and moderate - severe stutterers.

<table>
<thead>
<tr>
<th></th>
<th>M-group</th>
<th>MS-group</th>
<th>total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>PWR</td>
<td>19</td>
<td>28</td>
<td>47</td>
<td>34</td>
</tr>
<tr>
<td>AP</td>
<td>1</td>
<td>23</td>
<td>24</td>
<td>17</td>
</tr>
<tr>
<td>SP</td>
<td>2</td>
<td>22</td>
<td>24</td>
<td>17</td>
</tr>
<tr>
<td>WWR</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>PR</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>INT</td>
<td>3</td>
<td>20</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>R</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>BW</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>HP</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>total</td>
<td>36</td>
<td>104</td>
<td>140</td>
<td></td>
</tr>
</tbody>
</table>

Psycholinguistic variables

(1) Grammatical class. Most of the fluency failures occurred on content words (68 %). This finding was most striking for the severe stutterers. Mild stutterers seemed to have a more even distribution of disfluencies between content words and function words. Some studies have indicated that content words are stuttered upon more frequently than function words because they are more meaningful and important, because they have higher information value, and because they have a tendency to start with consonants.
(2) Word length. The three word lengths tested were monosyllabic (M), bisyllabic (B), and polysyllabic (P). The mean percentage of fluency failure was analyzed with respect to the occurrence of the words. Looking at the whole group, the results indicated that as syllabic length increases there is a corresponding increase in fluency failures. But, again, there is a difference in the pattern for the subgroups which is illustrated in Figure 2. The MS-group exhibited a significantly higher rate of disfluency on polysyllabic words than did the M-group.

Figure 2. Percentage of disfluency (%SyD) on monosyllabic (M), bisyllabic (B), and polysyllabic (P) words for mild (M) and moderate-severe stutterers (MS).
(3) Word position in the phrase. The first word, or words early in the sentence, are said to be stuttered upon more frequently than are the remaining words (Brown, 1938; Soderberg, 1967). Clauses analyzed in this study varied in length from 2 - 9 words (average 5.4 words). Table II gives a summary of the results. From the table it is clear that first words do not have the highest stuttering score. The first three words have only 47% of the total stuttering occurrences. This is contrary to a majority of the previous research, some of which was listed in the introduction to this paper. Rieber, Smith and Harris (1976) however, found in their study that the medial part of a sentence had a higher rate of fluency failures; which is in accordance with the present study. Soderberg (1967) has also found that stuttering tends to occur in "vollies" on odd numbered words (e.g. the 1st, 3rd, 5th, 7th). In this study, words number 3 and 5 have the highest percentage of stuttering. If disfluencies are added this effect become even more noticeable.

Table II. Frequencies of stuttering on words in clauses classified by length.

<table>
<thead>
<tr>
<th>No of words in Clause</th>
<th>Wordposition</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1</td>
<td>5</td>
<td>-</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>3</td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>-</td>
<td>4</td>
<td>2</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>totals</td>
<td></td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>14</td>
<td>19</td>
<td>6</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>96</td>
</tr>
<tr>
<td>%</td>
<td></td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>15</td>
<td>20</td>
<td>6</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

14
(4) Position within a word. The following positions were observed: stutterings on the initial sound (P1), the first syllable (P2), and medial syllable/s (P3). In Figure 3 the distribution of the total number of stutterings in percent are plotted for both groups. Initial sound and first syllable is without doubt the most conspicuous positions for fluency failures. Severe stutterers demonstrated far more stuttering on initial sounds, both repetitions and prolongations. They also had several fluency failures within a word, which was never demonstrated by the mild stutterers. The most common for this group was repetition of the first syllable, particularly repetition of monosyllabic words.

Figure 3. Distribution of the total number of stutterings in percent for mild (M), moderate-severe stutterers (MS). Initial sound (P1), first syllable (P2), medial syllable/s (P3).
A significant difference can be noted between stressed vs. unstressed syllables. Disfluency occurred on 29% of the stressed syllables compared to only 7% of the unstressed syllables (p<.05). It is possible that this explains the word position results, since in the speech material for this experiment accented words and stressed syllables did not have an early position in the sentence, in most cases. Thus, the fluency failures were most frequent in medial position which tended to be stressed.

With reference to phonetic characteristics the literature show conflicting data. Most confirmed stutterers would say that certain sounds are more difficult than others. Words beginning with a consonant are more difficult than words beginning with a vowel. Words beginning with a plosive are also said to be difficult. Although the present study is not controlled for phonetic balance, some general observations can be reported. Of the total frequency of initial plosives 29% were disfluent and most of these failures occurred on voiceless sounds. The amount of disfluent initial fricatives was 14% and of these the voiceless were in majority. Only 3% of initial vowels were involved in fluency failures. There was, however, considerable individual variability concerning "feared" sounds.

Perceptual judgment of disfluency

The second set of data dealt with listeners' perception of disfluency and their judgments on a five-point scale. The following measures were computed for each subject; points (p): listener ratings averaged for each phrase, and disfluency (SyD): percentage of disfluent syllables for each phrase.

In order to observe how naive listeners' ratings of fluency/disfluency correlated with "actual" disfluency measures, a linear regression analysis was made for each subject (Ferguson, 1972). The general equation is given by the formula y = kx + 1.
\[ y = \text{average points for each phrase} \]
\[ k = \text{slope of the line} \]
\[ x = \text{percentage of SyD for each phrase} \]
\[ l = \text{the intersection point on the y-axis} \]
\[ r = \text{Pearson product-moment correlation coefficient} \]

The greater the value of \( r \) the more accurate the prediction could be from one variable to the other.

For each subject the final scores, \% SyD, and computed linear regression values are illustrated in Table III.

Table III. Points (p), percentage of disfluent syllables (SyD) and values from linear regression analysis for each subject.

<table>
<thead>
<tr>
<th>Subject</th>
<th>p</th>
<th>%SyD</th>
<th>l</th>
<th>k</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4.38</td>
<td>5</td>
<td>4.83</td>
<td>-.091</td>
<td>-.94  *</td>
</tr>
<tr>
<td>6</td>
<td>4.20</td>
<td>7</td>
<td>4.79</td>
<td>-.078</td>
<td>-.85  *</td>
</tr>
<tr>
<td>1</td>
<td>3.95</td>
<td>8</td>
<td>4.66</td>
<td>-.082</td>
<td>-.95  *</td>
</tr>
<tr>
<td>5</td>
<td>3.53</td>
<td>16</td>
<td>4.51</td>
<td>-.063</td>
<td>-.82  *</td>
</tr>
<tr>
<td>4</td>
<td>2.82</td>
<td>25</td>
<td>4.01</td>
<td>-.047</td>
<td>-.70  **</td>
</tr>
<tr>
<td>2</td>
<td>2.78</td>
<td>20</td>
<td>3.69</td>
<td>-.042</td>
<td>-.76  *</td>
</tr>
<tr>
<td>1-6</td>
<td>3.60</td>
<td>14</td>
<td>4.55</td>
<td>-.068</td>
<td>-.87**</td>
</tr>
</tbody>
</table>

* \( p \leq .001 \) df = 13
** \( p = .01 \) df = 13
*** \( p = .001 \) df = 88

The results showed a highly significant negative correlation between listener judgment of disfluency and the observed rate of disfluency for each subject. The correlation coefficients are somewhat lower for subjects 5, 4 and 2. Interestingly these subjects constituted the subgroup moderate-severe stutterers (MS).

Inspection of Table III indicates that the more disfluent a subject is, the lower is the k-value. The more fluent the
subject is the steeper is the slope of the line and the higher the l-value. There is a strong relationship between percentage of disfluency and the points assigned.

The slope of the line has to do with the overall disfluency while stuttering severity is the dominant factor contributing to the l-value. It might seem self evident that listeners give lower points the more disfluent the speech is. It is, however, not quite so simple. Looking at each phrase and comparing phrases between subjects it becomes clear that the listeners take several factors into account when rating fluency.

One possible explanation for the low scores is the amount of disfluency itself in relation to the number of syllables. More important, however, is the nature of disruption. The listeners seemed to discriminate stuttering from normal nonfluency and stuttering events were more disturbing to the speech than were other forms of disfluency. This implies that words which lack integration between segments and/or syllables were considered to disrupt the speech more seriously than did word repetitions, phrase repetitions and revisions. Interjections were used too frequently by one subject (i.e. in some phrases prior to almost every word) and this kind of speech disruption also lowered the scores. As mentioned above, interjections will sometimes function as "starters". Silent prolongations seemed to be more disturbing than those that were vocalized.

Subjects 2 and 4 require special comment. They have similar average values both for points and k-values. Subject 4 has a mean of 25% SyD but only 13% SyS, while subject 2 has a mean of 20% SyD which all involved stuttering (20% SyS). Subject 2 got a lower value of 1 since this value is the more important in predicting stuttering severity.

As a final step, regression analysis was calculated for all of the 90 phrases with an obtained correlation coefficient of -.87 \( (r^2 = .76) \) indicating that 76% of the variance in the listeners responses had been explained.
Judgment of stuttering

The judges were also asked to indicate whether a phrase contained stuttering or not. For each phrase, a comparison was made between the listeners' and the experimenter's judgment. Of the 90 phrases there was total agreement on 41 phrases (46%). If the criteria was 75% of the listener panel there was 78% agreement between listeners' and experimenter's judgment of each phrase.

Measure of listener-rated overall severity of stuttering was estimated and correlated with measured mean percentage of stuttered syllables for each subject. The Pearson correlation coefficient was .997 (p < .001, df = 4), which means that stuttering severity for each subject was agreed upon to a very high degree. This high correlation is however deceptive, since an examination of each specific phrase showed greater variability in the judgment of stuttering. There was a tendency for more correct responses if the sentence contained no disfluency or more than one stuttering event. But the quality of stuttering (i.e. type and duration) seemed to be most important. If there was only one instance of a tense blocking or one instance of several repetitions of the same segment, the event was always labeled as stuttering by all observers.

A regression analysis based on all phrases was computed for the variables: experimenter judgment of stuttering / listener judgment of stuttering. This analysis revealed a significant correlation between the variables ( r = .80, p < .001, df = 88 ). The pattern that emerged when comparing the two groups of stutterers was the following: M-group: r = .82, p < .001, df = 43 and MS-group: r = .67, p < .001, df = 43.

No definition of stuttering was given prior to the listening session. The mere fact that the listener panel was asked to indicate "stuttering" on the response sheet might have had some influence on the outcome. Williams and Kent (1958) argue
that listeners tend to identify the same interruptions as stuttering when biased in that direction by the instructions, and as normal interruptions when told to listen for such.

DISCUSSION

Generalisations which can be drawn from the data presented in this article are somewhat limited given the relatively small number of subjects, but certain conclusions can be reached and some questions can be raised.

Young (1961) demonstrated that short segments of speech were sufficient to determine severity of stuttering and that randomly selected samples could be used. Furthermore, Williams et al. (1963) found that audio cues and audiovisual cues were in extremely close agreement for judging the frequency of stuttering. In the present experiment, an informal comparison of the utterances used as test samples with the total body of speech material from which they were selected, suggest a considerable agreement with the other researchers. Thus, the use of short segments of speech appears to be a valid technique and we may now turn to a discussion of the present data.

The results of the present investigation indicated that during the three speaking conditions the group of mild stutterers differed in systematic ways from the group of moderate-severe stutterers. There was a difference not only in frequency of disfluencies, but also with regard to type, loci and duration of stuttering events. It is obvious that frequency counts alone do not provide the only index of stuttering severity. The dissimilar group trends in frequency of fluency failures across the conditions are interesting but difficult to interpret. The severe stutterers were more disfluent in all conditions, but this was more prominent during the dialogue session. An obvious and rather dramatic downward trend was
apparent in the third condition. Langová and Šváb (1975) demonstrated, but did not explain, that stutterers were more disfluent in a listener's presence than when alone. In the present investigation this was true only for severe stutterers. Perhaps they were experiencing fear to a greater extent than were the mild stutterers. The dialogue situation required swift responses. The experimenter (who read the opposite lines) was not a stutterer, and this fact might have heightened the expectation of communicative disability.

One possible reason for increased fluency failures may be the prosodic factor. Stress and intonation patterns require co-ordination and exact timing and increased activity of the speech-mechanism. In Swedish there are two tonal accents, accent I and accent II. In accent II words, the presence of a main stress implies a secondary stressed syllable within the same word. The dialogue task contained more bi-and polysyllabic accent II-words and this had a greater effect on the speaking task.

The level of disfluency for mild stutterers was approximately the same as for severe stutterers in the third condition when situational anxiety was probably high for both groups. For the MS-group however, it was possible to avoid "hard" sounds/words by circumlocution and by using synonyms and fewer polysyllabic words. Stuttering, per se, decreased more than did disfluency. The subjects had to give specific information, which is said to increase disfluency, but the time pressure was less urgent than during the other conditions.

Clinicians and researchers are almost unanimous in the opinion that stutterers have trouble "getting started". That is, stuttering occurs more often on the first word or the first three words in the sentence. The present study, in agreement with several other, (e.g. Kaasin and Bjerkå, 1982; Rieber et al., 1976) show conflicting data regarding word position in a sentence. This suggests the need for further studies regarding the parameters of stress and intonation pat-
terns in relation to word position.

Most stutterers, mild or severe, state that they have certain sounds which are difficult to initiate especially in initial position of a word. In fact, they many times utter the sound itself successfully in its first part, but then prolong it or repeat the whole syllable. Thus, the breakdown doesn't always occur on the particular "feared" sound but later in the speech sequence. Sometimes stutterers deliberately repeat syllables in order to postpone the utterance of a feared sound. It is as if stutterers concentrate on saying sounds rather than words.

So, the question arises what are stutterers really doing when these disruptions occur within a word or when they block before a word? What are the basic mechanisms for the lack of integration and co-ordination of sounds, syllables and words?

In judging severity of disfluency there is an interplay between several variables such as type of disfluency, frequency of disfluency, duration of stuttering event, length of sentences, and perhaps voice quality, (e.g. one of the subjects had a dramatic shift in voice quality between the three conditions). Although there is a lack of data to carry this last argument further here, it presents an interesting question for future research.

It seems highly likely that not only form and frequency of stuttering but also duration of each event may guide the listener in search for cues to determine severity of stuttering. The listeners in this study were unsophisticated in the sense that they had no formal training regarding stuttering. This fact did not seem to affect the rating of overall severity which was in close agreement with the quantitative analysis. On the other hand, they showed more variability in correct discrimination between stuttering and normal non-fluency. Phrases containing more severe stuttering were more likely to be labeled as stuttering by all listeners. If the
listening task had been designed in another way (i.e. judging stuttering unit-by-unit in a phrase) the observer agreement would probably have been lower.

The results of the current study thus indicate that varying both speech material and speech tasks was an effective method for elucidation of various speech characteristics in audible speech behavior; both quantity and quality in speech disruptions were highlighted. In addition, the difference between mild and severe stutterers was demonstrated. It is evident from the data that reliance upon exclusively oral reading material or self-formulated speech is not a sufficient measure of disfluency in stutterers.

The method described in this study could be adapted to the clinic as a screening test of overt speech behavior. For a thorough assessment of the stutterer's difficulties it would need to be complemented with a case history and interviews, as the individual's own experience of his communication disorder is also important.

REFERENCES


INTELLIGIBILITY VS REDUNDANCY -- CONDITIONS OF DEPENDENCY

Sheri Hunnicutt

Abstract

The relationship between context redundancy and key-word intelligibility was examined in sentences having both high and low redundancy. The study partially replicates work by Philip Lieberman (1963) in which he concluded that the degree of stress and carefulness of articulation of a word varies inversely with its redundancy. More words and more extreme redundancies in word pairs were used in the current study. These word pairs were placed in similar positions in two sets of sentences: sentence pairs that one might find in text, and adages together with sentences that might be spoken. With the text-type sentence pairs, there was an intelligibility advantage for the words in lower redundancy contexts. For the adage and spoken-sentence pairs, there was no intelligibility advantage for words in either context. It is suggested that the adage is not a good representative of high redundancy contexts. The result that intelligibility and redundancy are inversely related in some instances (text-type sentences) indicates that information control by a speaker is going on even while reading test sentences. That is, utterance planning for the speech production mechanism involves interpreting semantic information which may be expressed in fine adjustments in motor control.
Introduction

It has been assumed, for some time now, that the intelligibility of a word out of context decreases as the redundancy of its context increases. That is, if the context of a word is quite redundant, the speaker does not expend much effort to make the word intelligible to his listeners since they can, for the most part, deduce it from context. The study which is most often cited for this assumption is one done in 1963 by Philip Lieberman. Lieberman concluded from his experiments that the degree of stress and carefulness of articulation of a word during speech production are approximately inversely proportional to the word’s redundancy. (The redundancy of a test word was established by having readers fill in the blank in the sentence it was removed from; the percent of correct guesses was the index of redundancy.)

A number of general questions are raised with respect to this conclusion. Do talkers naturally control for the amount of information they wish to give to a listener? Does an ideal speaker control his effort to make words within a message intelligible according to his perceived difficulty in having them understood by his listener? Are adjustments to this end made phonetically as well as grammatically and prosodically? And do these adjustments actually serve their purpose -- do they make speech more or less intelligible to the listener?

If the answers to these questions are affirmative, we can make some strong claims about the speech production mechanism. In particular, we can say that utterance planning includes very fine calibrations which are dependent upon the semantic relationship of each word to its sentential (or clause) context. That is, very high level information is being used to predict the realization of speech at the level
of formant frequencies, VOT, transition rates, relative intensities of phonemes within a word, and other low-level phenomena which have been shown to affect the intelligibility of a word (e.g., Chen, 1980 and Picheny, 1981).

The question we must ask now is the following: Is Lieberman's study general enough to allow such implications? In particular, can we assume that the material was representative? If the results are correct, the conclusion certainly holds for situations with a listener, and in situations in which the information transfer is even greater. Many such assumptions rest on the results which were reported. Turning to his experimental situation, we see that although a great deal of material was read by three speakers in Lieberman's study, only 11 sentences were used in the analysis, and only 7 words. The three words taken from the two adages had total-context redundancies of 0.6 and 0.85. The other four words, however, had total-context redundancies of 0.4, 0.2, 0.1 and 0.1 for the higher redundancy cases.

Because of the extensive use that has been made of these results, it seemed important to replicate the experiment with more words and with more extreme redundancies in the word pairs. It was hypothesized that little difference in intelligibility would be found for words in high and low redundancy contexts in a situation in which there is no listener, i.e., in a recording of speech read in an anechoic chamber. If any differences appeared, they were expected to be from the exaggeratedly redundant contexts, that is, in the words taken from adages. It was expected that these results would lead to further experiments in which sentences would be spoken for the purpose of communicating an idea to a listener and that only in such situations would differences in intelligibility appear.
Test Materials

In the study reported here, subjects attempted to identify words extracted from sentence contexts. The words were taken from similar high and low redundancy sentence contexts containing matched pairs of words. One set of high-redundancy sentences were adages. The corresponding set of low-redundancy contexts were from grammatically similar sentences that might be spoken. The two other corresponding sets of sentences were rather long, grammatically standardized sentences which one might find in a text. All sentences were read by one speaker.

The 76 text-type sentences were taken from a set of Swedish test sentences developed for speech perception tests by Rolf Lindgren (1982). These sentences have very similar grammatical structure; the test words are all common words beginning with a stop and are in sentence object position, far enough along in the sentence for a specific context to be built up. Redundancies for the sentences were established by having subjects fill in the blanks left by removing the test words.

The grammar of the sentences is as follows:

\[
\text{Adverbial - Verb - Subj - Adv. - Obj - }\begin{cases} 
\text{Relative Clause} \\
\text{Conjunction +} \\
\text{Short main clause}\end{cases}
\]
Several examples are given below with approximate English translations.

Redundancy

1.000 Hela natten läste studenten ivrigt boken som han fått låna.

(All night the student eagerly read the book which he had been able to borrow.)

.034 Nyfiket granskade mannen länge boken (breven) som han hade funnit.

(Curiously the man slowly examined the book (the letters) which he had found.)

.910 Under förmiddagen delade brevbäraren snabbt ut breven som samlats under helgen.

(During the morning the postman quickly delivered the letters which had collected during the weekend.)

Redundancy in the second sentence is low, .034. It is easy to see that there are many likely choices for the test word, e.g., the book, the letters, since many things can be slowly examined. We see that in the other two sentences, however, that choice narrows to include only the book in the first sentence and little more than the letters in the third, since what students read or postmen deliver is much more restricted, i.e., the redundancy is high.

The distinction between redundancy and probability can be observed in the third sentence. The sentence redundancy
is high -- it could be said that 91% of the essential sentence information is present without the test word. However, the probability of the particular word brev (letters) is low (6.7%) because the most expected word is post (mail).

The 36 Swedish adages were collected with the criterion that they contain a non-initial noun (preferably a two-syllable noun and preferably not sentence-final.) The 36 companion sentences were constructed to have similar grammatical and syllable structure as the adages with the test word in the same position -- in addition, the sentences should be typical spoken sentences. These sentences were chosen so as to give as little information as possible about the test word, i.e., they were constructed as low redundancy contexts. Two examples follow:

Nya kvastar sopar bäst.       (New brooms clean best.)

Långa kvastar kostar mer.     (Long brooms cost more.)

Man ska inte döma hunden efter håren.

(One shouldn’t judge a dog by his coat.)

Du ska inte skicka hunden efter barnen.

(You shouldn’t send the dog after the children.)

Half of each type of the sentences were used as actual test material, half as fillers.
Recording Conditions

One adult male Swedish speaker recorded the 148 sentences in randomized order. He was told to read each sentence once silently to himself, and then once aloud. This procedure was used to avoid any "surprises" while reading, i.e., the speaker knew the meaning of the sentence before reading it. The speaker sat about 30 cm. from a Bruel & Kjaer 1" 4145 microphone in an anechoic chamber. Recordings were made on a Telefunken (Magnetophon 28) tape recorder with a Bruel & Kjaer 2607 amplifier.

Speech Processing

The sentences were copied from tape onto a large computer disk. From the disk files, the test words were edited out using an interactive speech processing program (MIX, Rolf Carlson). Editing was first done by looking at the speech waveform and listening between markers placed in the waveform, and then adjusted by listening alone in the few cases in which a neighboring sound could also be heard before or after the word. A 2.1-second period of "pink" (speech-like) noise was edited in to cover each word. The noise level was chosen so that the average signal-to-noise ratio over the set of all words was 4 dB. The covering noise was used so as to give no cue to word length. The noise was increased from zero at the onset and attenuated at the offset over a period of 100 milliseconds to avoid an abrupt, and possibly distracting or confusing presentation. These words were then recorded on tape again with 4 seconds between each word and 10 seconds after each set of 15 words.

When the list of words was first heard by several native Swedish speakers, it was decided that the words were too easy to understand, and that the noise level should be increased.
(Plomp and Mimp en, 1979, report that an increase in the signal to noise ratio of 1 dB in the most sensitive range may be reflected in as much as a 15% to 20% increase in sentence intelligibility). To determine the appropriate noise level, the words were presented to 10 subjects with the assumption that a dB level could be established over which most answers would be correct, and under which most answers would be incorrect. This proved to be a faulty assumption -- dB level and intelligibility were not strongly correlated. A calculation was made to determine the correlation of the intensity with the number of correct answers. The correlation was .33. It was therefore decided not to carry out a new test with uniformly reduced intensities.

While making dB level readings on the test words, it was informally observed that the dB level seemed to reduce through a sentence, and that words taken from near the beginning/end of a sentence had comparatively high/low readings respectively. An analysis of this effect was made, and a correlation of .66 was found. In the figure on the next page (Figure 1), the baseline has been set at the level of the noise which was added in the experiment. It can easily be seen that the signal-to-noise ratio tends to reduce with sentence position, approaching zero at the end of a sentence. An additional vowel-specific analysis showed no particular dependence on vowel type.

The test results indicated, however, that there might be enough incorrect answers to analyze the data as had first been intended. It should be noted that these data are from only 10 subjects.
Figure 1: Signal-to-noise ratio as a function of the percentage of sentence (in syllables) preceding and including the stressed syllable of the test word
Results

For the 19 text-type sentence pairs, there was a clear intelligibility advantage for the words in lower redundancy contexts.

Number of sentence pairs in which the low redundancy words were more intelligible

10

Number of sentence pairs in which the low and high redundancy words were of the same intelligibility

7

Number of sentence pairs in which the high redundancy words were more intelligible

2

The mean number of correct answers for the words in low redundancy contexts was 8.79; for words in high redundancy contexts, 7.26. The means were tested with a paired comparison test (Box, Hunter and Hunter, p. 101) and were found to be significantly different at the 2.2% level. Neglecting the cases in which all answers were correct for both high and low redundancy contexts, the mean number of correct answers was 8.08 for words in low redundancy contexts and 5.67 for words in high redundancy contexts. These means were found to be different at a significance level of 1.9%. This suggests that a more difficult test would produce a clearer advantage for words in lower redundancy contexts.

What about the influence of a word's probability on its intelligibility, we might ask. The data here are few: there are 12 pairs with correlated redundancy and probability, and 7 pairs in which the high redundancy contexts are filled with low probability words. The mean number of correct answers for high probability words (implying a high redundancy con-
text, as well) is 6.92. For high redundancy contexts with low probability words, the mean number of correct answers is 7.86. The difference in the means was not found to be significant at the 10% level. (Unpaired design, Box, Hunter and Hunter, p. 115) Low probability words in high redundancy contexts were all equally or more intelligible than their high probability counterparts, however. That is, all deviations were in the direction of higher intelligibility for lower probability words.

For the 21 adage/spoken-type sentence pairs, however, there was no such advantage for words in lower redundancy contexts.

Number of sentence pairs in which the low redundancy words were more intelligible 5

Number of sentence pairs in which the low and high redundancy words were of the same intelligibility 8

Number of sentence pairs in which the high redundancy words were more intelligible 8

The mean number of correct answers was also quite close for the two groups, both with and without cases in which all answers were correct. The differences were not significant at the 10% level of confidence. Upon reflection, it appears that the context of an adage was not a particularly good choice for high redundancy. We are quite familiar with these "sayings," of course, but they are frequently uttered with the emphasis of making a point, not with the redundancy of
saying something everyone already knows. In addition, their metaphorical nature likely alters the source of meaning from the combined meanings of constituent words (or phrases) to that of the meaning of the message as a whole; this must strongly affect word production and intelligibility.

This finding, of course, does not agree with Lieberman's. For two of his three words from adages, this discrepancy may possibly be explained by the fact that these words were probably the recipients of sentence focus in the low redundancy case. They may thus have been more prominent than they would have been in a context similar to the adage.

Discussion

The results of the current study indicate that the relationship of intelligibility to redundancy is not clear. There may be a dependency in certain conditions, but not in others. The question that has been asked in this study, and also in Lieberman's study concerns the intelligibility of a word in isolation and its dependency upon factors of redundancy in context. That is, redundancy is defined as the percentage of essential information present in a sentence without the test word. Then we can say that in the non-idiomatic, non-metaphorical sentences of a reader, these results indicate that there is a clear intelligibility advantage for words in lower redundancy contexts. That is, the speaker compensates for the lower redundancy context so that the resulting word in that context is more intelligible in isolation. We may, therefore, assume that such information control goes on even when reading test sentences. This means that speech motor control, even in such cases where the intention to communicate is deficient, (i.e., where there is no listener), is modified by the degree of information intended.
For other sentences, such as the adages, no conclusions can be drawn from this study. We may conjecture that this effect is not always present because the intentions of the speaker vary, and because the essential information the speaker intends to convey may be expressed at various levels of abstraction.

Let us now consider another use of the word "redundancy" as the systematicity in one's language (and speech). This definition refers to the information in a complete sentence over and above that which is essential. How can these two uses of the word "redundancy" be compared? We have said that the redundancy which is an effect of language systematicity can be described as the information in a complete sentence over and above that which is essential. We could perhaps describe the redundancy calculated for the current study as the information in the sub-sentence, i.e., the sentence "minus" the test word, which is available to predict the missing word. One could also say that the redundancy in the sentence has been (possibly) reduced by omitting the test word.

Other research on redundancy in speech has frequently been concerned with the ability to use this systematicity in language in certain conditions. Accessibility to stimuli from which to make systematic inferences may depend upon environment and manner of speaking. In the presence of noise or a manner of speaking that degrades the speech, one would expect information (and redundancy) to be decreased. In such a situation, a listener would not be able to take advantage of a language’s systematicity to the same extent.

One might wonder then, if the resulting speech shows that the speaker can compensate for the reduced redundancy, and whether the intelligibility of the speech can be restored on account of this compensation? And, at the other extreme,
if redundancy is facilitated by attempting to speak clearly in a noise-free environment, is intelligibility improved for a somehow "deficient" listener?

One answer for the case in which redundancy is reduced comes from a study by J.C.T. Ringeling. He first defines two types of redundancy reduction, one from outside causes such as noise, which he calls "external reduction," and one from degraded speech, called "internal reduction." He chooses one type of internal reduction, i.e., speaking in a low, or soft voice, and asks whether the speaker successfully restores the intelligibility of his speech under such a condition. His hypothesis was that soft voice speech would be more intelligible than normal voice speech if both versions were attenuated to the same barely audible level, i.e., that the low voice speaker had compensated successfully for the lower intensity of his speech. This, in fact, was not the case. He concludes that no form of phonetic compensation was introduced by the speaker for sentences spoken in a soft voice.

So, now, what about "clear speech" and other strategies to be understood above noise or by someone who has difficulty in hearing or in comprehending? Is greater intelligibility achieved if redundancy is facilitated by such a strategy? Some contributions in this area come from two theses at M.I.T. by Picheny and Chen. In these studies, it was shown that, for most speakers, speaking clearly resulted in significantly more intelligible speech for both the normal-hearing (Chen) and listeners with sensorineural hearing losses (Picheny). The results were verified at various signal-to-noise ratios (Chen) and over all phoneme classes (Picheny).

We are led to say, then, that even though speakers seem to be able to make fine distinctions in motor control for the purpose of making specific words in variously redundant
contexts intelligible, that redundancy reduced by manners of speaking or the effect of outside noise cannot be equally well compensated for. It might be useful, then, to distinguish between situations in which redundancy is globally reduced, and those in which redundancy varies dependent on context. In the latter situation, a speaker may well be able to monitor the listener's ability to understand, and be able to judge the threshold of intelligibility, to good effect; in the former situation, the speaker may not succeed in his effort, especially in the case of internal reduction of redundancy such as with "soft voice" or whispering. In the case of external reduction of redundancy, recognition of this failure may lead to various non-normal speech modes such as shouting above noise or "speaking clearly" to the hard-of-hearing or to foreigners or language learners. Some of these speech modes, "speaking clearly," for example, may be successful.

It appears that the answer to our question is at least partially dependent upon speech mode. But speech mode is, in turn, dependent upon the speaker's possibly complex intentions. That is, social factors play an important role here. In order to "factor out" these social influences for further studies, it would be instructive to examine both what is known about the phonetic correlates of different speech modes, and what is known about vehicles of intelligibility.

Conclusions

This study has partially replicated a study done in 1963 by Philip Lieberman in which he concluded that the degree of stress and carefulness of articulation of a word varies inversely with its redundancy. Lieberman's study has been cited extensively by speech researchers, and many assumptions about information control by speakers have been made which
depend upon the validity of the results. Because of the importance of these results, and the nature of the experimental material, it was decided that these experiments should be replicated with more words and with more extreme redundancies in the word pairs.

Two sets of sentences were used in this study, sentence pairs that one might find in text, and adages together with sentences that might be spoken. With the text-type sentence pairs, there was an intelligibility advantage both for the words in lower redundancy contexts and for words of lower probability. For the adage and spoken-sentence pairs, there was no intelligibility advantage for words in either low-redundancy (spoken-type sentence) or high-redundancy (adage) contexts. It was conjectured that the metaphorical nature of adages and the influence of the social conditions in which they are used may have prevented the realization of the intelligibility-redundancy relationship observed in the text-type sentences.

Situations in which redundancy is reduced by noise or a manner of speaking that degrades the speech, as well as situations in which redundancy is facilitated by strategies such as "clear speech" were also discussed. It was noted that speakers are not always able to compensate for this type of reduced redundancy. That is, lower redundancy does not, in general, imply increased intelligibility due to speaker compensation.

It is concluded that one cannot make a general assumption that lower redundancy results in increased intelligibility. However, it is clear that in many cases, there is an intelligibility advantage for words in lower redundancy contexts. This indicates that information control by a speaker is going on even while reading test sentences. That is, utterance planning for the speech production
mechanism involves interpreting high-level semantic information in fine adjustments to motor control.

References


THE ROLE OF VOWEL CONTEXT IN THE PERCEPTION OF PLACE
OF ARTICULATION FOR STOPS

DIANA KRULL

Abstract

In recognition algorithms and certain theories of speech perception the interpretation of the signal is based on "distance scores" for comparisons of the signal with stored references. The present study consists of listening tests whose main purpose was to gather perceptual data for a calibration of such distance metrics by investigating the perceptual confusions of the Swedish stops [b,d,q,g'] in various systematically produced fragments of $V_1C:V_2$ that provided 25 vowel contexts for each consonant. An additional goal was to assess the importance of vowel context for the identification of intervocalic stops from the point of view of the invariance problem.

The results showed that most listeners need the vowel context in order to identify the consonants, the following vowel being more important than the preceding one. Some subjects obtained very high scores on stimuli with only 26 ms following stop release which points to the existence of release cues of some consistency.

The confusions in the perception of the 26-ms stimuli were shown to form a systematic pattern dependent on the original vowel context. For example, often [g'] from a front vowel context was perceived as dental or retroflex, while [g] from a back vowel context was perceived as labial.
1. INTRODUCTION

A. Background

(i) Different views on place cues for stops

Many years of investigation on cues for place of articulation in stops have led to a variety of conclusions about the nature of these cues. The opinions differ, especially, with regard to what kind of cues are to be considered as the primary carriers of place information, and whether these cues are to be considered as invariant or context dependent.

According to one view, the essential place cues are dynamic and context dependent. They are to be found in the time-varying formant transition pattern mirroring the articulatory gesture after stop release (Liberman, Cooper, Schankweiler, and Studdert-Kennedy, 1967; Liberman and Studdert-Kennedy, 1978).

Dorman, Studdert-Kennedy, and Raphael (1977) also regard place cues as being context dependent, but argue that essential place information is carried by the transitions in some vowel contexts, while in others the burst is more important.

A different claim has been made by Blumstein and Stevens (1979). They contend that primary cues to stop place of articulation are invariant and reside in the gross spectrum shape integrated over a short time following stop release, no matter whether the spectrum contains burst, transitions, or
A view similar to that of Blumstein and Stevens is held by Kewley-Port (1983), but whereas Blumstein and Stevens treat onset spectra as static, Kewley-Port stresses the importance of their dynamic aspect.

While most investigators have concentrated on some specific cue, Schouten and Pols (1983) have studied the relative importance of a variety of cues. Using Dutch material with the consonants varied in both position and vowel context, Schouten and Pols demonstrated that both bursts and transitions contain a great deal of information in any vowel context, with the burst being the more important one wherever there is a significant difference. They also demonstrated that various aspects (e.g. burst, transition, stationary vowel) taken together contain more information than would be predicted by a probabilistic model on the basis of the information in each one of them taken alone.

(ii) Is acoustic invariance absolute or relative?

Blumstein and Stevens (1979) hypothesize absolute invariance on the acoustic level. In more recent studies, however, acoustic invariance is regarded as being based on changes in the relative distribution of energy in the burst, rather than on some absolute qualities (Lahiri and Blumstein, 1981; Ohde and Stevens, 1983).

In a different sense, acoustic invariance is also relative rather than absolute when seen from the point of view of distinctive feature theory as outlined by Jakobson, Fant,
and Halle (1952). As Fant has pointed out (Fant, 1973, Chapter 10), the acoustic correlate of any feature will vary with the simultaneous, preceding, and following features, and with prosodic elements.

The question thus arises: is absolute invariance necessary? Perhaps, as Lindblom (1984) suggests, a minimal condition would be sufficient contrast so that semantically distinct information can be presented in perceptually distinct form.

(iii) Distance metrics as part of the description of speech perception

In several recent investigations, distance metrics have been used for the description of speech perception. The method is also considered a useful tool determining rules for automatic speech recognition on the auditory level. It has been used to describe vowel perception by Bladon and Lindblom (1981), Carlson and Granström (1979), Carlson, Granström, and Klatt (1979), and Klatt (1982).

The method is based on the idea of a perceptual space whose coordinates are chosen among auditory parameters such as, for example, loudness, pitch, timbre, etc. The distance between stimuli in this space is measured in psychoacoustic units, for example, just noticeable differences, or in units on a subjective judgement scale.

In connection with the present work, the papers by Schouten and Pols (1979a,b; 1981) are of special interest. They have investigated transition movements in Dutch CVC words.
(six initial consonants, four vowels, and five final consonants), and have shown that for most of the consonants there is a common locus area in a perceptual space for the transitions of the same consonant through different vowel contexts. The locus area is more clearly defined for CV than for VC transitions.

B. Purposes of this paper

This paper is the first in a three-part investigation and consists of perception tests with the immediate aim of answering the following questions:

(1) How well can Swedish listeners identify intervocalic voiced stops when the vowel context, or part of it, is removed?

(2) When listeners fail to identify a consonant, what do they hear instead?

(3) Do possible confusions form some systematic pattern that depends on the (original) vowel context.

There are two reasons why answers to these questions are important in the context of the preceding discussion. One reason is our interest in the invariance issue. Using intervocalic Swedish stops may provide some information of interest in this connection, especially since there are four places of articulation for Swedish stops. Moreover,
intervocalic stops have been less frequently investigated from the invariance point of view than have initial or final stops.

The second, and primary, reason for posing the above questions is that we want to use the resulting perceptual data for a calibration of distance metrics. We propose to do this using an algorithm that takes acoustic signals, in this case the test stimuli, as input and gives as output the location of each stimulus in a perceptual space in relation to a stored reference. The assumption used as a basis for this is that if a stimulus is correctly identified then it is located at the least distance from the reference. The goal, given an acoustic description, is to predict the number of confusions. A method of making such predictions is a kind of perception model and is needed for both practical and theoretical reasons. It is needed, for example, in automatic speech recognition and for an objective evaluation of deviant speech.

The actual calibration of distance metrics will constitute the third part of this investigation, while the second part will consist of acoustic measurements.

2. PROCEDURES

A. Stimuli

The initial stimuli consisted of nonsense words of the form \( \hat{V}_1C:\hat{V}_2 \) read by a male Swedish phonetician, native speaker of the Middle Swedish dialect. The consonants \([b,d,q,g]\) were preceded and followed by the vowels \([\acute{e},\epsilon,a,\backslash] \) or \([\Upsilon] \). The
words were read using the Swedish "grave" accent in which the pitch is falling on the first syllable and rising on the second and both syllables have about equal prominence.

All of the one hundred possible combinations were read from a randomly ordered list with the subject seated in an anechoic chamber.

The tape-recorded material was subsequently digitized with a 20 kHz sampling frequency (low pass 8 kHz) and stored in a computer. From these initial stimuli, referred to henceforth as "Whole word", an additional three tapes were prepared by cutting the speech wave with the help of a computer program written by Rolf Carlson, Royal Institute of Technology, Stockholm. The first cut was made at the zero crossing nearest to the middle of the consonantal closure. The resulting stimuli were called "First half" and "Second half". Next, a cut was made at the zero crossing just before the consonantal release, and an additional cut at the zero crossing 26 ms further on in the speech wave.

The 26 ms duration was chosen because it was the window width used by Stevens and Blumstein (1978) and Blumstein and Stevens (1979). They gave the following argument for using just this duration: "A 26-ms window seemed to give spectra for which the the gross attributes associated with the different places of articulation were most salient, although these spectral attributes were not strongly dependent on window length." (Stevens and Blumstein, 1978, p. 1360). Stevens and Blumstein used a Hamming window while the window in the present experiment was rectangular.

Using the same stimulus length for all four places of articulation involved certain problems. As is well known, the
length of the noise burst varies with place of articulation. Thus in this experiment labial stimuli often include part of the following vowel while for velars part of the noise burst may be cut away.

An alternative would have been to use different lengths for the stimuli, for example always including one glottal pulse of the vowel. In that case, however, the stimuli would have varied in length according to place of articulation, thus providing an additional cue to consonant identity.

The final stimuli were transferred to four tapes:

1. Whole word 100 stimuli 1 occurrence each
2. First half 300 " 3 occurrences each
3. Second half 300 " 3 " "
4. Burst 300 " 3 " "

On each tape, the order of the stimuli was randomized. At the beginning of each of the first three tapes 20 extra stimuli were added, 10 for practice and 10 as a buffer. To the "Burst" tape 40 extra stimuli were added, 20 for practice and 20 as a buffer.

The interstimulus interval in all tapes was 4 sec except after each group of ten stimuli where the interval was 10 sec.

B. Subjects

24 native speakers of Middle Swedish dialect served as subjects. Seven of the 24 were phonetically naive and about 17 years of age. They were paid for their participation. The
remaining 17 subjects were volunteers from among the staff and students of the Institute of Linguistics, they ranged in ages from 22 to 57 and had varying degrees of phonetic experience.

The data from four of the subjects were analyzed separately for two reasons.
(1) An audiometer test showed that two of the subjects had hearing losses.
(2) Two subjects, although with normal hearing, had much lower correct scores in "Whole word" stimuli than the rest of the listeners (79% and 81% while the rest of the subjects had scored 90% or more). These wrong identifications involved dental and retroflex stimuli only.

C. Testing procedures

The tapes were presented through earphones in a quiet room, in two listening sessions on different days. Day one: "Whole word" and "Second half"; day two: "First half and Burst". The tapes were thus ordered according to increasing difficulty. The subjects were allowed a few minutes rest between tapes and in the middle of each tape.

The listening level was determined adjusting a 1 kHz calibration tone so that it "covered" the stimulus with the highest level. Before the tape was presented to the listeners, the calibration tone was set to 89 dB SPL ± 1 dB, the variation being due to differences in earphone calibration. In this way the "Burst" stimuli were presented at a level 4 dB higher relative to the level at which they occurred in "Whole word". The difference was not larger, contrary to
what could have been expected, because the labial "Burst" stimuli often contained part of the following vowel which caused their level to be relatively high.

The listeners' task was to identify the consonant and, at the same time, indicate how confident they were of their identification by writing a number from 1 to 5 in their preprepared answer sheets, "1" for "Not at all confident", "5" for "Completely confident", and the numbers in between for intermediate degrees of confidence. The listeners were asked to give an answer to each stimulus even if they had to guess.

3. RESULTS

A. Number of correct identifications for each place of articulation with comments on inter-subject variation and learning effect

(i) Scores for 20 subject group

Percent correct identifications for each consonant and stimulus range ("Whole word", "First half", "Second half", and "Burst") are shown in Fig. 1. The diagrams show average data and ranges of variation for the 20 subjects. It can be seen that apart from the variation due to differences in the degree of difficulty in different stimulus ranges, there is a large variation both across places of articulation and across
Fig. 1. Percent correct answers per consonant for 20 subjects in four stimulus ranges. (● mean value, —— range between highest and lowest single values.)
Labials were well identified in all stimulus ranges with a relatively small inter-subject variation.

Velars were even better identified than were labials in the first three stimulus ranges, with almost no inter-subject variation. In "Burst" stimuli however, the correct scores for velars were considerably lower and the inter-subject variation larger than for other places of articulation.

Scores for retroflex were low in all stimulus ranges, and were only slightly better for dentals. However, counting dental and retroflex together as one class, the answers were 100% correct for all 20 subjects in "Whole word" and "Second half", while for "First half" and "Burst" the correct score was lower but comparable to that of labials, and much higher than the separate scores for dental and retroflex.

In view of the great inter-subject variation, t-tests were carried out to see if there was any significant difference between the two groups of subjects, the first consisting of six naive subjects, the second of fourteen listeners with varying degree of phonetic experience. The tests showed that the retroflex scores for the naive listeners were significantly lower than those of the "sophisticated" group in all stimulus ranges (p<0.01 for "Whole word" and "Second half", p<0.02 for "First half" and "Burst"). There was no significant difference between the groups for either labials or velars, and for dentals there was a significant difference in "First half" only (p<0.05); here, as in retroflex, it was the correct score for the naive listeners that was lower.

Although retroflex identification seems to have been
easier for listeners with phonetic experience, it was nevertheless more difficult than the identification of other places of articulation, at least in all stimulus ranges except "Burst".

All scores were also checked for possible learning effects and such an effect was found to be present in the scores of the "Burst" stimuli. Therefore, eight of the twenty subjects were asked to listen to these stimuli a second time. The results showed an increase in percent correct from 73 in the first session (for these eight subjects) to 79 in the second session. The difference is significant at the 1% level. Although each one of the eight subjects did better the second time, there was a large difference in the amount of increase in correct answers with the lowest being 2% and the highest 17%.

The inter-subject variation in percent correct was also greater in the second session, ranging from 59% to 93%. Although there was no significant learning effect in the second session for the eight subjects seen as a group, two of the subjects performed increasingly better all through the test. In the last 50 stimuli one of these subjects had only three wrong answers and the other two.

Although there thus appears to be a quantitative difference between the results of the two listening sessions, a preliminary analysis of the confusions has shown that there is no qualitative difference. The confusions in both sessions are of the same kind and occur in the same vowel contexts. The confusions in the first session are more salient, though, both because a greater number of mistakes were made per listener and because there were more listeners. We therefore decided to
use the results of the first session for a more detailed analysis.

(ii) Scores for the subjects whose results were analyzed separately

The results of two subjects, N1 and S2, were analyzed separately because of their unusually low scores in the "Whole word" stimuli, 79% correct for S2 and 81% for N1 (see Table I). For both these subjects, all confusions in "Whole word" as well as in "Second half" were between dental and retroflex. In "First half" the difference between percent correct for these subjects was greater. Moreover, while S2 confused mainly dental and retroflex even here, N1 had unusually low scores for labials as well. As can be seen in Table I, the difference between percent correct for these subjects was still greater in the "Burst" stimuli. Both had low retroflex scores here, but while the other places of articulation were above group mean for S2, for N1 these too were low, especially the score for [g] which was only 11% correct. Both these subjects had, as has been remarked earlier, normal hearing. It is therefore difficult to explain the low scores and the differences between the two subjects.

Two other subjects had their results analyzed separately because, as noted above, they had demonstrated hearing losses on standard audiometric testing. Subject S7 had a loss of slight to medium severity in high frequencies in both ears, and subject S11 had a monaural loss of medium severity in all frequencies tested. The scores of these subjects were within
the limits of one standard deviation of the group, except the "First half" score for S7 which was lower.

<table>
<thead>
<tr>
<th></th>
<th>N1</th>
<th>S2</th>
<th>S7</th>
<th>S11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole word</td>
<td>81</td>
<td>79</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td>Second half</td>
<td>76</td>
<td>74</td>
<td>97</td>
<td>99</td>
</tr>
<tr>
<td>First half</td>
<td>76</td>
<td>84</td>
<td>64</td>
<td>93</td>
</tr>
<tr>
<td>Burst</td>
<td>36</td>
<td>64</td>
<td>64</td>
<td>63</td>
</tr>
</tbody>
</table>

Table I. Percent correct answers for the subjects analyzed separately

B. Consonant confusion analysis

(i) Confusions across vowel contexts

Confusion matrices for each stimulus range for the pooled vowel contexts can be seen in Fig. 2. In "Whole word" and "Second half", the confusions were almost exclusively between dental and retroflex. Confusions between these two places of articulation were also high in "First half", but in this case the listeners also had some difficulty with labials and, to a lesser degree, with velars.

Correct scores for "Burst" stimuli were lower for all consonants and especially for [g] which was confused with each of the other three consonants in about an equal degree. Dental and retroflex were as in other stimulus ranges confused mostly with each other.
Fig. 2. Confusion matrices for 20 subjects in four stimulus ranges. Vowel contexts pooled.
(ii) The role of vowel context

There were few "Whole word" and "Second half" confusions and these were limited to dental-retroflex; it is therefore difficult to say anything conclusive about the influence of vowel context in these cases.

Fig. 3. shows the effect of the preceding and following vowel on the identification of "First half" stimuli. It can be seen that the effect of the following vowel is small. The slight lowering of correct scores when the following vowel was [ɛ] or [ɔ] was largely due to mistakes with the retroflex, but also to some extent with labial recognition.

The effect of the preceding vowel was greater. The low percent correct after [I] was for the most part due to the fact that that retroflex was perceived as dental, and, to a lesser extent, to confusions of labial with dental. The dip at [a] is primarily due to numerous confusions of retroflex with [g] for the stimulus ['agːɛ].

As might be expected, the effect of the vowel context was strongest for the "Burst" stimuli. Complete confusion matrices for each of the 25 vowel contexts are given in Fig. 4. In each row of matrices in Fig. 4 the preceding vowel changes from front to back while for each column of matrices it is the following vowel that changes in the same manner. (No result is given for the stimulus [ɔbːɛ] because two of its samples were faulty. The third was identified 100% correctly by all subjects.)

Comparing results by vowel contexts and consonants, it can be seen that the confusions form a regular pattern. For example, [g] in front vowel context was often confused with
Following vowel Preceding vowel

Fig. 3. Percent correct answers in "first half" (VC) stimuli with a) different following vowel; b) different preceding vowel. All consonants pooled.
### Fig. 4. Confusion matrices for "Burst" stimuli in 25 vowel contexts. See text for details.
dental and retroflex, but seldom with labial. In back vowel context, on the other hand, [g] was often confused with labial but almost never with dental or retroflex. The consonants seem to have been easiest to identify in the context of [a].

For a clearer illustration of the influence of both the preceding and the following vowel, plots were drawn giving the different vowels on the horizontal axis and percent identifications as [b,d,d] or [g] on the vertical axis. Fig. 5 shows the influence of the preceding vowel and Fig. 6 that of the following one. A comparison between the two figures shows that the influence of the following vowel was usually the strongest. For [g] there was a curious difference in the effect of the vowels: with a preceding front vowel the correct score for [g] was lower than with a back vowel in the same position while for following vowels the opposite was true: the correct score was higher in front vowel context. However, for both preceding and following vowel it was the case that back vowels gave more confusions with labials.

Note also the influence of the vowel context, especially the following vowel, on dental and retroflex identification. As the vowel changed from front to back, identifications as [d] decreased and listeners heard more often [q]. However, there was a difference between the respective influence of the two back vowels [ɔ] and [U]. With [ɔ], especially in final position, listeners gave a [d] response more often than when the vowel was [U].
Fig. 5. Consonant identification in "Burst stimuli with different preceding vowel."
Fig. 6. Consonant identification in "burst" stimuli with different following vowel.

Following vowel: Ω - e - a - c - Ω

Stimulus: b

Percent answers
(ii) Degree of confidence

The subjects' degree of confidence ("1" for lowest and "5" for highest) varied with stimulus range. Thus "Whole word" and "Second half" had the highest mean degree of confidence, "First half slightly lower, and "Burst lowest (see Table II).

<table>
<thead>
<tr>
<th></th>
<th>Whole word</th>
<th>First half</th>
<th>Second half</th>
<th>Burst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean degree of confidence</td>
<td>4.7</td>
<td>3.5</td>
<td>4.7</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Correlation:

number of answers - confidence

|            | 0.64       | 0.62       | 0.53        | 0.62  |

Table II. Mean degree of confidence and correlation coefficients between the number of answers given per place of articulation for each stimulus and the mean degree of confidence for these answers.

The correlation between the number of answers in each square of the confusion matrix in Fig. 4 and the corresponding mean degrees of confidence are also shown in Table II. The low correlation coefficient for "Second half" is due to high degrees of confidence being given with most of the answers.

The eight subjects who listened twice to the "Burst" stimuli were more confident of their answers in the second session where they gave a mean degree of confidence of 2.1 which is considerably higher compared to their 1.7 rating in the first session.
Within the same stimulus range, correct answers usually had higher degree of confidence. There were exceptions, though, in two of which the listeners were quite consistent. One of these cases was retroflex perceived as dental. The degree of confidence in this case was slightly higher than for the correctly identified retroflex consonants in almost all vowel contexts. The other exception was velar identified as labial in back vowel context where the degree of confidence was as high as for the corresponding correct identifications.

In "Burst" stimuli, listeners were consistently more sure of their answers when identifying a consonant as [b], less sure of identifications as [g], and least sure of identifications as [d] or [q].

4: DISCUSSION AND CONCLUSIONS

A. Why were retroflex consonants so difficult to identify?
Comments on the phonological status of retroflex in Swedish

The results reported above show that the greatest number of confusions occur between dental and retroflex. Even in the case of "Wole word" stimuli, only 25% of the subjects were able to identify all dentals and retroflexes correctly. The confusions between dental and retroflex in the shorter stimulus ranges can therefore not be explained by the missing vowel context only. It may, at least in part, be due to "top down" processing and thus have its explanation in the
ambiguous phonological status of the retroflex consonants in Swedish.

The ambiguity results because the retroflex consonants [t, q, t'] can be considered to be phonemes if seen from the point of view of the consonant system, but not when seen from the point of view of their distribution. On the one hand there are, for example, numerous minimal pairs such as

\begin{verbatim}
karl [kɑːl] - kal [kɑːl]
barn [bɑːn] - ban(a) [bɑːn]
sport [spɔtː] - spott [spɔtː]
bord [buːd] - bod [buːd]
\end{verbatim}

On the other hand, supradentalization occurs also over word and morpheme boundaries, e.g. /d/ in du becomes retroflex if preceded by /r/, thus bor du [buːdər], and the /r/ itself is subsequently deleted. Similarly, the ending -na becomes [ŋa] in e.g. korna [kuːŋa], etc. In these cases the retroflex clearly is a derivation of /r/ + a dental phoneme.

The view currently favored in phonology is that supradentalization is a two-step process: an assimilation of dentals to the alveolar place of articulation followed by a deletion of /r/ (Linell, Svensson, and Öhman, 1971).

One explanation for the difficulty listeners had in distinguishing between dental and retroflex may thus have been that they felt both to be in some sense the same. (The retroflex stop was described by one of the subjects as "a kind of thick d".)

An additional difficulty in this experiment may have been caused by the phonological inadmissibility of segment
combinations in some of the stimuli. There were three kinds of inadmissible combinations. Firstly, [q] does not normally appear in utterance initial position as it seems to do in "Second half" and "Burst" stimuli; it does, however, appear in stressed syllable-initial position, e.g. gardin, gardera, pardon [paqɔn:], [paqun:], kardan. Secondly, [q] very rarely appears after non-low front vowels - this may partly explain the especially large number of retroflexes identified as dentals in "First half" after [l]. Thirdly, [q] does not normally appear after a short vowel (except when it is a result of assimilation over word boundary, in which case the consonant is short too). In recent years, however, forms like [bʊqːɛ] and [juqːɛ] have begun to appear in the Stockholm dialect.

Other investigators have also reported difficulties with retroflex. Öhman (1966), for example, used initial and final segments of /aCCa/ stimuli in his perception tests. The correct scores for /rda/ were low, but not for for /ard/, in spite of the vowel being short.

The difficulties with retroflex identification include non-auditory stimuli as well. Traunmueller (1974) carried out perception experiments with Swedish subjects using a visual lip-reading aid (no acoustic information), the correct scores for place of articulation were lowest for retroflex.
B. The role of the length of the noise burst relative to the length of the "Burst" stimuli

As is well known, the length of the noise burst varies with place of articulation. Labials, for example, have a short burst, sometimes no burst at all, while velars have the longest burst. Since stimulus length is the same for all places of articulation, labial stimuli contain part of the vowel (most often about two glottal pulses) while velars sometimes have part of the noise burst cut away. The fact that the correct scores for "Burst" stimuli are highest for labials and lowest for velars suggests a relation between burst length and stimulus length. Specifically, the stimulus length of 26 ms may be too short for good identification of velars. This would agree with the observation of Stevens and Blumstein (1980) that while 10-20 ms is sufficient to identify labials and alveolars, the auditory system may need longer time to build up a representation of a compact onset spectrum. Another possibility is that listeners require the beginnings of the transitions for their identifications.

C. Can the regularities in the confusion patterns be explained on the basis of distinctive feature theory?

The results for "Burst" stimuli have shown that the confusion patterns are clearly dependent on the original vowel context of the stimuli. The manner in which confusions vary with vowel context suggests a possible relationship, on the
distinctive feature level, between the the consonant the
listeners thought they heard and the adjacent vowel.

To determine whether there actually is such a
relationship, the features of the perceived consonant were
compared with those of the adjacent vowel(s). The distinctive
feature system used is that proposed by Fant (1973) for
Swedish. It is in its essentials based on the system of

According to this system, there are three pairs of
features that are shared by vowels and consonants in this
experiment: (1) compact - diffuse; (2) acute - grave; (3) flat
- plain.

Although the same features are used for both vowels and
consonants this does not reflect actual identity.
Nevertheless, from the abstract acoustic point of view, there
are similarities in the interrelations of (ptk) and (uia),
(see Fant, 1973, Chapter 10).

In the case of this experiment, [b] is distinguished from
[d] and [g] by the same features as [U] is distinguished from
[I] and [a]. The same is valid for the distinctions [d] and
[g] on the one hand and [I] and [a] on the other. (See Table
III.)

The feature flat which distinguishes retroflex consonant
from dental is not needed here to distinguish any of the
vowels from one another, because there are no rounded front
vowels included, and the back vowels are distinguished from
each other by the compactness feature.
Table III. Distinctive features (from Fant, 1973) shared by the vowels and consonants of this experiment

What confusions might be expected on basis of the data in Table III, assuming that if the consonant cannot be identified correctly the listeners would tend to hear the consonant that shares essential features with the adjacent vowel? In that case listeners ought to hear \([d]\) if the vowel context has been \([I]\), \([b]\) if the vowel context has been \([U]\), and \([g]\) if it had been \([a]\).

Examination of the results for stimulus \([g]\), where most of the confusions occurred (Fig. 4-6), shows that this is precisely what happened, especially if both the preceding and the following vowel are taken into account. Not only was \([g]\) in \([U]\) context perceived as \([b]\) but listeners were also very confident of their identification (see 3. B. (iii)). In \([\gamma]\) context the effect was still there but weaker. In \([I]\) context listeners heard the acute and diffuse consonants \([d]\) or, to a lesser degree, \([q]\), which suggests that \([-\text{compact}]\) and \([+\text{acute}]\) may have been more important than the flatness feature in this case.

The role of the flatness feature cannot be assessed here
with any certainty, partly because no rounded front vowels have been used in the stimuli, and partly because [+flat] for consonants has been defined as "A shift down in the frequency location of formants retaining the general shape of the spectrum" (Fant, 1973, Chapter 8). Here, a shift down may have been effected also by an adjacent grave, i.e. [-acute] vowel. That more [d] stimuli have been judged as retroflex in back vowel context could thus have been predicted, as could be the greater number of judgements of [d] stimuli as dental in front vowel context. What is difficult to explain, however, is the greater number of retroflex identifications in the context of [\textipa{ fø }] when compared to the context of [\textipa{ u }]. (This was true for both retroflex and dental stimuli.) The feature that distinguishes between [\textipa{ fø }] and [\textipa{ u }] is that of compactness. Since [\textipa{ fø }] is classified as [+compact] we would expect less retroflex identifications here when compared to the [\textipa{ u }] context, especially since, in addition to this, [\textipa{ fø }] can be classified as [+flat] and [\textipa{ u }] as [+flat].

Confusions of [\textipa{ b }] stimuli are too few to allow any conclusions to be drawn.

It is interesting, though, that confusions with [\textipa{ g }] are more frequent in front vowel context when the stimulus is [\textipa{ b }], but in back vowel context when the stimulus is [\textipa{ d }] or [\textipa{ q }]. The explanation to this may be that in both cases the energy maximum of the stimulus is displaced towards the middle range, thus conveying the effect of compactness.

Forthcoming acoustic measurements will show how much the the consonant spectra have, in fact been affected by the vowel context.
D. Can we learn something about possible invariance of the place cues from the results of the perception tests with the 26-ms stimuli?

The mean correct scores for "Burst" stimuli are much lower compared to those of the "Second half", which suggests that most listeners need the information in the following vowel in order to identify the consonant.

This does not, however, exclude the possibility of invariant cues. One reason for this is the high percentage of correct identifications for some single subjects (see 3. A. (i)). There must have been some cues that they could base their identifications on. What these cues are can not be said without acoustic measurements. An additional reason supporting the possibility of invariant cues is the learning effect (see 3. A. (i)), especially the fact that not all subjects reached a plateau in their correct scores, but got better as the test went on. It is therefore impossible to say how good they might have been at the identification of the "Burst" stimuli given some more training.

Although there obviously are place cues in the 26-ms samples, not all listeners have used them. That different people can use different cues to identify stops, and do so quite consistently, has been shown by Walley and Carell (1983).
E. Conclusions

The perception tests have shown that vowel context is important for consonant identification in two ways.

First, there is a decrease in the number of correct identifications when the vowel context, or part of it, is removed. The following vowel is more important for the identification than the preceding one.

Second, and more important, the confusions in the perception of "Burst" stimuli form a systematic structure that is clearly dependent on vowel context and may be explained with the help of distinctive feature theory.

On the basis of the results of the perception tests the possibility of invariant cues can not be ruled out.

As was mentioned in the Introduction, however, this paper is only the first of a three-part investigation. It will be followed by acoustic measurements and, subsequently, by a calibration of distance metrics. The conclusions drawn in this first part must therefore be regarded as preliminary.
LITERATURE


Vowel categorization by the bilingual listener

Richard Schulman

ABSTRACT

A study was made examining categorical perception for Swedish listeners highly fluent in American English. The investigation consisted of two listening sessions whose major difference was not in the stimuli, but in the formulation and language of instruction. The stimuli were 23 synthetic vowels along a continuum from [I] to [a] appearing in a "s_t" frame for the English tape and "s_tt" frame for the Swedish one. It was hypothesized that the categorization of the stimuli would be different for the two language conditions, reflecting the differences existing between the two languages' vowel systems. The results show category boundary shifts for [I] - [a] and [ɛ] [a]. A marked increase in response inconsistency at the English [ɛ]- [a] boundary as compared to that for the Swedish was also observed. It was concluded, supporting findings by Elman et al, that a change in task instructions can be sufficient for eliciting different perceptual criteria in a categorization task.

1 The research reported here has been supported by the Swedish Council for Planning and Coordination of Research
INTRODUCTION

It has long been known that a listener's identification of the same speech sounds may vary depending on the context within which they are presented. This was early demonstrated by an investigation by Ladefoged and Broadbent (1957), in which listeners tended to assign a different phonological quality to the same vowel of a test word, this in accordance with the vowel characteristics of the carrier phrase containing the word. We conclude, then, that the listener can adjust his perceptual criteria for a specific speech unit dependent on the speech characteristics of the speaker at hand.

In a study demonstrating across-listener variability in the identification of identical speech sounds, Lisker and Abramson (1964) showed that language background was influential in determining whether a listener would categorize stops as voiced or voiceless. A stop with a specific VOT (voice onset time) was classified as voiceless by speakers of one language, while it was classified as voiced by speakers of another language. Such results seem reasonable considering that the languages in question use different VOT's in the production of the voiced/voiceless distinction, and that these distinctions in production might lead to the listeners developing different perceptual criteria.

The results of the Ladefoged and Broadbent study suggest that criteria used for the identification of speech sounds are influenced by the listener's evaluation of the speech which a speaker has already produced plus the expectations concerning the speech which the speaker will produce. From the Lisker
and Abramson investigation we see that the phonetic differences of the sounds in the phonological inventories of languages (or dialects) will cause phoneme categorization to differ depending on the language base of the listener. From these results we can hypothesize that, if a person is fluent in two languages he will apply different perceptual criteria to the same speech stimuli when these stimuli are presented to him in two different language contexts.

In the present study we investigate this hypothesis. We make the basic assumption that the bilingual listener has two sets of criteria - one for the phonetic distinctions of each language. However, unlike experiments which rely on varying the contrast between the test stimulus and its context to bring about a change in the listener's identification criteria (Ladefoged and Broadbent 1957; Sharf and Ohde 1981, adaptation and VOT; Simon and Studdert-Kennedy 1978, adaptation and anchoring in experiments using /b/-/d/ continuum; Summerfield 1979, timing contrasts and VOT), we attempt to influence the listener to access one set of criteria or the other solely by changing the language of presentation and instruction for the test.

Using a similar approach, Elman et al. (1977) investigated the shift in categorization of /b/ and /p/ by English-Spanish bilinguals. Presenting naturally produced stops differing in VOT, they found that of their subjects, those judged to be most bilingual exhibited switching of phoneme categories depending on the language of presentation. This is strong evidence in favor of our hypothesis. However, when they retested these subjects with similar stimuli (though this time synthetically produced), the shifts in categorizations were not as extensive as those with the natural stimuli, supporting results by Caramazza et al. (1973) who also used synthetic stimuli and
found no apparent category switching. It was suggested that other acoustic cues (such as the burst and the change in F0) which differ in the natural speech productions for the voiced/voiceless distinction in stop consonants, together with and not just VOT alone, were instrumental in causing the observed category shifts for the natural stimuli. This implies that it is not the case of natural or synthetic per se which is the deciding factor for shift vs. non-shift, but rather whether or not enough of the essential acoustic parameters are manipulated in the creation of the synthetic stimuli for them to elicit responses identical to those of corresponding natural stimuli.

In the present experiment, the task was to categorize synthetically produced vowels. We assume that the determinants of vowel quality in English and Swedish (the two languages used in the investigation) are primarily the formant frequencies of the vowel plus their relation to the fundamental frequency, and that other acoustic variables are of minor significance for vowel identification. Then, by varying only the formant patterns, we would expect to observe a category shift similar to that found in the Elman et al. investigation with natural stimuli. The vowel categories to be investigated here are the short vowels [I], [ɛ], [æ] and [a]. We can present two hypotheses concerning differences in the responses to the English task as compared to the Swedish. First, the [I] of American English is in general lower than the corresponding standard Swedish vowel (Schulman and Lindblom forthcoming) so, if there is a shift in categorization, we would expect it to be in that direction. Secondly, in standard Swedish, the phonemes /e/ and /ɛ/ merge when realized phonetically as the short vowel [ɛ]. The short vowel [æ] is also an allophone of /ɛ/ and, as a rule, only exists before a retroflex. In an earlier
study using the same stimuli as in the present one (Janson and Schulman 1983), it was found that Stockholm listeners could not distinguish well between the two phonemes. It was found, however, that there will be a slight prominence of [æ] responses between [ɛ] and [a]*. Since [æ] is an established phoneme in standard American English, we might expect the English stimuli to receive a larger region of [æ] responses than the Swedish stimuli, displacing the [ɛ]-[æ] border towards [I] and the [æ]-[a] border towards [a].

1. PROCEDURE

The experiment consisted of two listening sessions. In the first session the test was administered by an American, and in the second session by a Swede. When the English tape was played, the task was to identify each stimulus as either "sit" ([sIt]), "set" ([sEt]), "sat" ([sæt]) or "sot" ([sat]). The choices for the Swedish task were "sitt" ([sIt]), "sett" ([sEt]), "sätt" ([sEt]) and "satt" ([sat]). Each stimulus consisted of the frame st taken from a natural male utterance of the Swedish word "sett" with one of 23 synthetic vowels inserted medially. The stimuli were made with an OVE III synthesizer at the Royal School of Technology, Stockholm. The end point vowels had as the first three formant frequencies (steady state values): F1=325 Hz, F2=2200 Hz, F3=2850 Hz for [I]; and F1=760 Hz, F2=1275 Hz, F3=2550 Hz for [a]. These endpoint values were converted to Mel, and the three dimensional Euclidean distance between them was then

*In that study, the phonetic representations: [I], [æ], [ɛ] and [a] were used as opposed to the present study's usage of [I], [ɛ], [æ] and [a].
calculated. This was in turn divided into 23 equidistant 30 Mel steps (Table I).

<table>
<thead>
<tr>
<th>Vowel</th>
<th>F1</th>
<th>M1</th>
<th>F2</th>
<th>M2</th>
<th>F3</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>325</td>
<td>406</td>
<td>2200</td>
<td>1678</td>
<td>2850</td>
<td>1945</td>
</tr>
<tr>
<td>2</td>
<td>347</td>
<td>430</td>
<td>2165</td>
<td>1662</td>
<td>2824</td>
<td>1935</td>
</tr>
<tr>
<td>3</td>
<td>370</td>
<td>454</td>
<td>2130</td>
<td>1646</td>
<td>2800</td>
<td>1926</td>
</tr>
<tr>
<td>4</td>
<td>393</td>
<td>478</td>
<td>2093</td>
<td>1629</td>
<td>2774</td>
<td>1916</td>
</tr>
<tr>
<td>5</td>
<td>417</td>
<td>503</td>
<td>2059</td>
<td>1613</td>
<td>2750</td>
<td>1907</td>
</tr>
<tr>
<td>6</td>
<td>441</td>
<td>503</td>
<td>2025</td>
<td>1597</td>
<td>2724</td>
<td>1897</td>
</tr>
<tr>
<td>7</td>
<td>465</td>
<td>551</td>
<td>1992</td>
<td>1581</td>
<td>2699</td>
<td>1887</td>
</tr>
<tr>
<td>8</td>
<td>490</td>
<td>575</td>
<td>1957</td>
<td>1564</td>
<td>2676</td>
<td>1878</td>
</tr>
<tr>
<td>9</td>
<td>515</td>
<td>599</td>
<td>1925</td>
<td>1548</td>
<td>2650</td>
<td>1868</td>
</tr>
<tr>
<td>10</td>
<td>532</td>
<td>615</td>
<td>1872</td>
<td>1522</td>
<td>2643</td>
<td>1865</td>
</tr>
<tr>
<td>11</td>
<td>548</td>
<td>630</td>
<td>1821</td>
<td>1496</td>
<td>2635</td>
<td>1862</td>
</tr>
<tr>
<td>12</td>
<td>565</td>
<td>646</td>
<td>1770</td>
<td>1470</td>
<td>2628</td>
<td>1859</td>
</tr>
<tr>
<td>13</td>
<td>581</td>
<td>661</td>
<td>1721</td>
<td>1444</td>
<td>2623</td>
<td>1857</td>
</tr>
<tr>
<td>14</td>
<td>598</td>
<td>677</td>
<td>1672</td>
<td>1418</td>
<td>2615</td>
<td>1854</td>
</tr>
<tr>
<td>15</td>
<td>616</td>
<td>692</td>
<td>1624</td>
<td>1392</td>
<td>2608</td>
<td>1851</td>
</tr>
<tr>
<td>16</td>
<td>634</td>
<td>708</td>
<td>1578</td>
<td>1366</td>
<td>2600</td>
<td>1848</td>
</tr>
<tr>
<td>17</td>
<td>651</td>
<td>723</td>
<td>1533</td>
<td>1341</td>
<td>2593</td>
<td>1845</td>
</tr>
<tr>
<td>18</td>
<td>669</td>
<td>739</td>
<td>1488</td>
<td>1315</td>
<td>2585</td>
<td>1842</td>
</tr>
<tr>
<td>19</td>
<td>686</td>
<td>754</td>
<td>1444</td>
<td>1289</td>
<td>2578</td>
<td>1839</td>
</tr>
<tr>
<td>20</td>
<td>705</td>
<td>770</td>
<td>1400</td>
<td>1263</td>
<td>2573</td>
<td>1837</td>
</tr>
<tr>
<td>21</td>
<td>723</td>
<td>785</td>
<td>1357</td>
<td>1237</td>
<td>2565</td>
<td>1834</td>
</tr>
<tr>
<td>22</td>
<td>742</td>
<td>801</td>
<td>1315</td>
<td>1211</td>
<td>2558</td>
<td>1831</td>
</tr>
<tr>
<td>23</td>
<td>760</td>
<td>816</td>
<td>1275</td>
<td>1186</td>
<td>2550</td>
<td>1828</td>
</tr>
</tbody>
</table>

Table 1
Formant values of stimuli vowels, in Hz and in Mel
For synthesis of the stimuli, the Mel values for each stimulus were converted back to Hz (for a more detailed description of the rationale used, see Janson and Schulman 1983). The locus values for the vowel transitions were 1950 Hz for F2 and 2900 Hz for F3. The formants above F3 were held constant for all stimuli. The duration of each vowel was 160 msec and the fundamental frequency would begin at 102 Hz rising to 115 Hz. The only difference between the stimuli for the two sessions was the length of the burst for the /t/. In standard American English, /t/ in word final position is only slightly released, while in standard Swedish it is both stronger and longer. To make the English stimuli, copies of the Swedish stimuli were made whence the burst of the /t/ for each word was cut away after 18 msec (as compared to the original Swedish burst of 130 msec.), with its amplitude left unchanged. This reduction of the /t/ burst should not influence the perception of the vowel, but only add to the impression that the words actually were English. The editing procedure was performed on an Eclipse S-200 computer at the Phonetics Laboratory, Stockholm University. The tapes consisted of six lists of the 23 stimuli, with each list being randomized internally. The order of the 6 lists was different for the two test tapes though within each list it was the same. The inter-stimuli interval was 4 sec except after every tenth word where it was 8 sec. There was no special pause between the lists.

The subjects were 10 female students at the University of Stockholm, 9 of whom were studying at the English department and had passed that department’s qualifying examinations certifying them as fluent speakers of American English. Of the 10 subjects, 8 had lived in the U.S., one in Canada and one near a U.S. military installation in West Germany, all for at least six months and before the age of 20. Informal conversations
with each subject revealed, however, that the extent of fluency varied from person to person.

The subjects were tested first with the English tape, and then with the Swedish, with at least a two week period between the sessions. The tapes were played through headphones at a comfortable listening level. The written instructions for the English test were as follows:

In this task you will decide if the words which you hear sound most like sit, set, sat, or sot. You are to write, in each space, the word which you hear. You are to answer each time, even if you are unsure. Write down the answers as soon as possible. The words come quite closely after each other with a longer break after every tenth word. Remember, this is a task to which there are no right or wrong responses. What we are interested in is how you identify each word.

For the Swedish test, the instructions were translated as closely from the English as possible. As can be seen from the instructions, the subjects were told to respond by writing down the whole word, so as to encourage listening for gestalts and not only for vowels. Before each test, the examiner engaged the subjects in conversation in the language of the session. It was also stated, before the second session, that though there might be similarities between the two test tapes, the tapes were not the same.

II. Results

In Fig. 1a and 1b are plotted the responses to the English and Swedish test tapes, respectively, for one representative subject. This subject's responses typify the tendencies shown
by the group as a whole for the two test situations. The first list for each test was counted as a training set and thus discarded, leaving 5 lists of the 23 stimuli. Along the x-axis on the top is the stimulus vowel number, and on the bottom each stimulus' distance in Mel from the border between stimulus No.1 and a hypothetical stimulus preceding it. On this axis, then, 15 Mel corresponds to stimulus No.1, 345 Mel to stimulus No.12, 525 Mel to stimulus No.18. Along the y-axis are the number of identifications for each stimulus. The responses are represented here orthographically. For the English task, we see a clear example of categorical perception for all the vowels.

Fig. 1a Responses by subject No.9 for the English task
For the Swedish task, however, we see this only at the borders of "i/e" and "ä/a". Between these two borders, the "e" and "ä" responses are rather mixed. As we have mentioned before, in production, most speakers of Central Swedish do not have [æ] except before a retro-flex, thus, "sätt" is pronounced [sɛt] as is also "sett". It is therefore understandable that the listeners' responses between the "i" and "a" categories will be a mixture of the vowels "e" and "ä", approaching 50% for each. Note, however, that there is clearly a preference for "e" at the border to "i", and for "ä" at the border to "a". For this subject, we do see that in agreement with our expectations for the Swedish responses as compared to the English, the boundary for [ɪ] against [ɛ] has shifted towards stimulus No.1, and
the boundary of [a] against [æ] has shifted in the same direction.

Next, we consider the responses of the subjects as a group for the Swedish tape as compared to the English tape. In our attempt to do this, we will apply a computation measure described by Janson (1979) called the "proportion mean" from which we can calculate the position of a category boundary in Mel. Table II gives the different vowel boundaries' computed values for each subject plus each boundary's region of uncertainty.

<table>
<thead>
<tr>
<th>Subject No.</th>
<th>i against eao</th>
<th>Position in MEL</th>
<th>Region of Uncertainty</th>
<th>No. of Stimuli</th>
<th>ie against aõ</th>
<th>Position in MEL</th>
<th>Region of Uncertainty</th>
<th>No. of Stimuli</th>
<th>ieå against ìo</th>
<th>Position in MEL</th>
<th>Region of Uncertainty</th>
<th>No. of Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>303</td>
<td>4</td>
<td>483</td>
<td>3</td>
<td>561</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>243</td>
<td>1</td>
<td>429</td>
<td>4</td>
<td>555</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>255</td>
<td>0</td>
<td>393</td>
<td>7</td>
<td>585</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>279</td>
<td>3</td>
<td>363</td>
<td>3</td>
<td>501</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>279</td>
<td>2</td>
<td>513</td>
<td>3</td>
<td>573</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>231</td>
<td>12</td>
<td>447</td>
<td>1</td>
<td>609</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>279</td>
<td>12</td>
<td>429</td>
<td>4</td>
<td>573</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>231</td>
<td>2</td>
<td>429</td>
<td>7</td>
<td>555</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>249</td>
<td>2</td>
<td>465</td>
<td>3</td>
<td>561</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject No.</th>
<th>i against eãå</th>
<th>Position in MEL</th>
<th>Region of Uncertainty</th>
<th>No. of Stimuli</th>
<th>ie against åo</th>
<th>Position in MEL</th>
<th>Region of Uncertainty</th>
<th>No. of Stimuli</th>
<th>ieâ against æo</th>
<th>Position in MEL</th>
<th>Region of Uncertainty</th>
<th>No. of Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>261</td>
<td>1</td>
<td>429</td>
<td>9</td>
<td>525</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>237</td>
<td>3</td>
<td>417</td>
<td>10</td>
<td>513</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>255</td>
<td>3</td>
<td>393</td>
<td>4</td>
<td>495</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>237</td>
<td>2</td>
<td>381</td>
<td>8</td>
<td>525</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>237</td>
<td>2</td>
<td>471</td>
<td>11</td>
<td>543</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>273</td>
<td>2</td>
<td>441</td>
<td>8</td>
<td>483</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>261</td>
<td>3</td>
<td>483</td>
<td>3</td>
<td>507</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>225</td>
<td>2</td>
<td>375</td>
<td>6</td>
<td>525</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>183</td>
<td>3</td>
<td>363</td>
<td>10</td>
<td>483</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>249</td>
<td>1</td>
<td>411</td>
<td>8</td>
<td>489</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table II  Vowel boundary position for each subject and language situation
Each subject is designated by the same number for both the English and Swedish test conditions. The region of uncertainty for a particular vowel is the region (in number of stimuli) in which responses for not only that vowel but also other vowels have been elicited. Looking then at Fig. 1a (subject No. 9 in Table II), the region of uncertainty for "i" against "eao" is 2, (that is, for stimuli numbers 8 and 9 both "i" and "e" responses are elicited), for "ie" against "ao" is 4 and for "iea" against "o" is 7. Using these values, we see presented in Table III, the mean position in Mel and the mean region of uncertainty for the group for each vowel boundary and condition. Also shown, is the significance of the difference between the two conditions as computed using a one-tailed T-test for paired values. In our analysis of the results, we will accept a level of p<.05 or less as significant. A significant difference in uncertainty implies that, comparatively, the subjects were consistent in their identifications for one test condition and inconsistent in their responses for the other. If the vowel boundaries (expressed as position in Mel) differ significantly from each other for the two tests, we may conclude that a boundary shift has occurred.

<table>
<thead>
<tr>
<th>Position in MEL</th>
<th>Region of Uncertainty No. of Stimuli</th>
<th>Region of Uncertainty No. of Stimuli</th>
<th>Region of Uncertainty No. of Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>i against eao</td>
<td>ie against ao</td>
<td>iea against o</td>
</tr>
<tr>
<td></td>
<td>i against eao</td>
<td>ie against ao</td>
<td>iea against o</td>
</tr>
<tr>
<td>Means English</td>
<td>262</td>
<td>438</td>
<td>563</td>
</tr>
<tr>
<td></td>
<td>242</td>
<td>416</td>
<td>509</td>
</tr>
<tr>
<td>Means Swedish</td>
<td>242</td>
<td>416</td>
<td>509</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>NS</td>
<td>.005</td>
</tr>
<tr>
<td>Significance Level p&lt;</td>
<td>.05</td>
<td>NS</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>NS</td>
<td>.005</td>
</tr>
</tbody>
</table>

NS=not significant

Table III  Comparison of mean responses for English and Swedish tasks
Looking at Table II, we see that for "i" against "eao" there is not a significant difference for the region of uncertainty between the two tests. There is, however, a significant difference between the two boundaries. This can be interpreted as a shift in vowel categorization in the expected direction. Looking next at the boundary between "ie" and "ao" (Eng.) against "ie" and "ää" (Swe.) we find that the difference in uncertainty is significant. That the subjects were inconsistent in their Swedish "e/ää" designations is not surprising. The 15 subjects from Stockholm tested by Janson and Schulman (1983) responded similarly, having a mean of 7.1 for this boundary's region of uncertainty. This relatively large uncertainty reflects a rather diffuse boundary between "e/ää", making a comparison between this boundary and the English one for "e/a" rather meaningless.

For the third vowel category boundary, the difference in regions of uncertainty is very significant. Of the three category boundaries for the English test, the subjects displayed the highest uncertainty (5.6) at this one, contrasting quite strongly with their identifications of the Swedish stimuli for which their identifications of [a] showed the greatest certainty (1.7). Once again, this uncertainty makes it difficult to compare the two boundaries. However, the significance between the two conditions' boundary positions is so large that it should not be ignored. The shift in boundaries, it can be observed, has occurred in the predicted direction.
III. DISCUSSION

The original intent of this study was to see to what extent a person fluent in two languages can access, perceptually, different phonetic criteria for either language. The power of suggestion was an integral part of the procedure. The subjects were told that the words they would hear belonged to one of the two languages, though in reality they belonged to both languages. The listeners, then, convinced that they were hearing a particular language, reacted in a manner appropriate for that language.

Similar attempts have been made before (Caramazza et al. 1973, Elman et al. 1977), but not successfully with synthetic stimuli. Elman et al. do concede that it might have been the lack of control of the appropriate acoustic parameters in the creation of their synthetic stimuli which resulted in a failure to replicate the results of their experiment with natural speech stimuli. Considering that it is in general much easier to synthesize natural sounding vowels than natural sounding stop consonants, this conclusion is indeed credible.

The results of the present study were somewhat unexpected. We had predicted differences in vowel categorization depending on the language context of the test. The differences, we hypothesized, would be expressed as the shifting of vowel boundaries. Shifts in vowel categorization were observed at the boundary between [I] and [ɛ] and also between [æ] and [a]. These, however, were not the only manifestations of an apparent difference in perceptual criteria for the two language conditions. We were also able to observe a distinct difference in the ability to consistently identify stimuli within a particular category. The responses for the Swedish "e/å" were too inconsistent to discuss, even in general terms, a
comparison with the corresponding English vowels. Similarly, there was a significant increase of uncertainty for the identification of the English word containing [a] ("sot") as compared to its Swedish counterpart ("sat"). That the subjects did not always succeed in identifying a stimulus consistently for one language condition while being able to do so for the other clearly supports the "different language context - different perceptual criteria" argument.

The mixture of "e" and "ä" responses for the Swedish task was, as mentioned previously, an expected result. But, how can we explain the apparent difficulty for the subjects to draw a sharp boundary between a and its neighboring vowel for the English task? There are several possible explanations, both in terms of language competency and phonetic reality. It was previously mentioned that there appeared to be different degrees of fluency among the subjects. These differences in ability may also exist in perception. We are not, however, suggesting that there is a strict correspondence between production and perception, for the possibility exists that a person's abilities for one can be more developed than for the other. Perceptual fluency is a topic which has, in general, escaped the thrust of language research and can certainly benefit from studies of this type. No matter how fluent the subjects for this experiment might be in English, that they only lived in an English speaking environment for a relatively short period of their lives implies that their command of English cannot equal their command of Swedish - hence the greater uncertainty. We have support for this argument in the results of Elman et al., who upon testing with the natural speech stimuli found the most category shifting by the most bilingual subjects. Had we, in our study, used subjects bilingual from childhood, we most probably would have found
lesser uncertainty for these vowels plus a more clearly defined category shift.

As for [a] in "sot", it is possible that it is not the unfamiliarity with the vowel which produced uncertain identification, but instead unfamiliarity with the word itself, "sot" being not a very common word. Upon having read the English instructions, all of the subjects professed ignorance of the word and asked to be told its meaning. If it is true that the uncertainty for [a] as in "sot" is due to the unfamiliarity with just that word, this has implications towards the use of nonsense words as opposed to real words in perception experiments. "Sot", in this case, could be considered a nonsense word - it being unknown to the subjects prior to the English test session. Moreover, the identification of [a] in American English does not have a history which is problem-free. When Peterson and Barney (1952) plotted the frequency of unanimous listener agreement for the identification of nine vowels, [a] received the least agreement. Compounding the situation in the present study, F2 for the [a]-like vowels does not go below 1275 Hz, which is a considerable difference from the average average F2 of 1090 Hz for [a] in the Peterson and Barney study.

The point which should be made is, regardless of whether one looks at the [I] - [ɛ] shift or the increased uncertainty for the other English vowels, it is demonstrated that listeners use different perceptual criteria under different speech conditions. The question then arises, why different perceptual criteria? In vowel production, we can point to two main factors which account for inter-speaker variation. One is anatomical, the other is dialectal. If we, as listeners, use some sort of perceptual scaling (Fant 1975, Traunmuller 1983) to normalize for the size of the speaker's vocal tract, we are
left with an idealized representation of the vowel. This representation, however, must still be normalized for its dialectal components. It is to this end that we have different perceptual criteria for different phonological systems. Anatomical scaling factors will tell us who the speaker is, phonological criteria will then tell us what he is saying.
REFERENCES


Lisker, L. and Abramson, A.S. (1964) Across-language study of voicing in initial stops: acoustical measurements. Word 20, 329-422


COMPREHENSION OF FOREIGN ACCENTS

(a Cryptic investigation)

Richard Schulman and Maria Wingstedt

The research reported here has been supported by the Swedish Council for Planning and Coordination of Research (Forskningsrådsnämnden).
The main purpose of this investigation is to study the strategies a listener employs when trying to understand a speaker with a heavy foreign accent. Since the study is, at this point, far from completion, our primary goal here will be to describe the methods we have used, rather than emphasize the results we have obtained.

A foreign accent can be said to be the consequence of a speaker’s application of the phonological rules of a language (usually the mother tongue) upon the target language. If the phonological systems of the two languages are very different, one may expect the foreign language learner to have difficulty in acquiring the idiomatic pronunciation of the target language. The relative success in learning this pronunciation varies from individual to individual, but rare is the person who does not have a trace of the mother tongue in his second language articulation.

Certain pronunciation errors, it seems, are more disturbing for the listener than others when trying to understand what the non-native speaker is saying. With systematic training, these pronunciation errors may be diminished and the task of the listener becomes easier. However, such training is not always available or particularly successful. If this is the case, it is the listener whose effort must increase in order to enable successful communication to take place. The role of the listener, then, becomes that of a more active interpreter of the speech signal.

The questions we ask, then, are: By what means does the listener accomplish this? What are the strategies one employs to decode a pronunciation highly deviant from the idiomatic one?

One hypothesis is that the listener, when exposed to the foreign accent, forms general phonological interpretation rules roughly corresponding to the phonological production rules which gave rise to the accent. Another hypothesis is that the listener uses a gestalt
strategy, whereby specific words are stored in an "equivalence lexicon". In this case, a particular pronunciation is identified as being equivalent the intended word, but no phonological analyses or generalizations are made to extend this to other words. Our assumption is that both processes are real, though interpretation by rule formation is the primary means for comprehension.

This assumption raises further questions: How much exposure to the foreign accent is needed before the listener can form such interpretation rules? Are certain error types more difficult to compensate for than others? How do linguistic aspects, other than the phonological, influence comprehension? To seek answers to these questions, we planned the following investigation, the target language being Swedish.

It was our original intention to find a non-native speaker with a heavy accent and record this person reading a Swedish text. Native speakers of Swedish would then be asked to listen to the recording and, after every sentence, repeat what they thought had been said. An analysis would then be performed, comparing the listeners' responses with both the original text and the speaker's production. However, we soon realized that this procedure would not be without complications. Even if we were to find a speaker with an accent so strong that comprehension would at times be impossible, it would be unfeasible to expect that such a speaker would be able to read aloud a Swedish text fluently. In addition, we could not control that the accent would be both unchanged and systematic throughout the reading. Further, it might be difficult to find subjects who had not been previously exposed to the accent in question.

Our solution to these problems was to create the foreign accent ourselves. We would compose a "cryptic accent", one which we could control by selecting specific phonological rules and applying them to
the Swedish text. The product would be a systematic, and therefore, predictable accent.

This "cryptic" phonological system is based on an inventory of common errors by non-native speakers of Swedish as reported by Bannert 1980. We have selected typical errors of pronunciation primarily from the foreign accents of Spanish, Greek and Turkish speakers. We consider the resulting phonological system to be a possible system, one giving rise to a conceivable foreign accent when applied to Swedish. The phonological rules are the following:

1. Quantity (and consequently also quality) distinctions are neutralized in consonants as well as vowels.

   "tak" [tə:k] \( \rightarrow \) [tak]  "roof"

   "tack" [tæ:k] \( \rightarrow \) [tak]  "thanks"

2. A neutral vowel is inserted in initial consonant clusters beginning with a stop.

   "kvar" [kva:r] \( \rightarrow \) [ką,var]  "remaining"

   "platt" [plæt] \( \rightarrow \) [pɔlət]  "flat"

3a. The stressed vowels of Swedish are changed into a simple five vowel system consisting only of unrounded front vowels and rounded back vowels.

   \( y \rightarrow \text{i} \)

   \( e, \varepsilon \rightarrow \text{ç} \)

   \( a \rightarrow \text{a} \)

   \( \theta, \alpha, o \rightarrow \text{c} \)

   \( \omega, \varepsilon \rightarrow \text{u} \)

3b. Unstressed vowels are reduced to a neutral [ə]-vowel.
4a. Voice assimilation: a voiceless consonant becomes voiced when followed by a voiced consonant.

"bort från" [bot fəːn] → [bord vron] "away from"

"svordom" [svuːdom] → [zvordam] "swear word"

4b. /s/ becomes voiced in intervocalic position.

"resa" [zeːsa] → [teːzə] "travel"

5. Intervocalic voiced stops become fricatives.

"grabbär" [gəːbːær] → [gəraːr] "guys"

"sedan" [seːdan] → [stɔːn] "then"

"vägen" [vɛːɡɛn] → [vɛːɡɛn] "the road"

6. Voiceless stops are unaspirated.

"på" [pɔː] → [pɔ] "on"

7. Supradentalization does not occur, and /r/ is pronounced as a tremulant retroflex.

"fors" [fɔːs] → [fɔːr] "stream"

8. [ʃ, ç, ʃ] → [s]

"sjuk" [ʃuːk] → [suk] "sick"

"tjäna" [ɻɛːna] → [sɛnə] "earn"

9. /h/ → [x]

"han hade" [hanː hadːɛ] → [xan xaðə] "he had"

10. /ŋ/ → [ŋɔ]

"ringa" [ɻɪŋa] → [rɪŋə] "to call"
11. Word stress is always placed on the penultimate (i.e. the second to last syllable). This rule is always applied before the rule for vowel insertion (2). The Swedish "grave word accent" (accent 2) is never applied, only "acute accent" (accent 1).

"arbete" [áːbɛːːtɛ] → [árbɛːtø] "work"

A pilot study was made in which these phonological rules were applied to a coherent Swedish text, 30 short sentences from a novel by Lars Molin (1968). Several sentences were modified so as not to exceed a length of 11 words. The result was a phonetic transcription of a cryptic version of the text. Sentences and not single words or phrases were used in order to present the listener with a task resembling a natural speech situation, in this case, listening to a story being read. An additional advantage of using a coherent text, is that it allows the listener to utilize the cues for comprehension which semantic and syntactic context might give.

One of the authors (R.S.) trained reading the phonetic transcription of the phonologically adapted text, the result being a Cryptic accent. A recording was made in which the Cryptic speaker read each sentence individually and with pauses in between. In the ensuing test tape, each sentence was followed by a 10 second pause. The task of the subjects was to listen to each sentence and repeat as literally as possible, in the pause following the sentence, what they had comprehended. The subjects were explicitly told not to imitate the foreign accent.

The subjects were tested individually and in isolation in an anechoic chamber. This avoided any sort of visual feedback from the test leaders. The tape was heard through Sennheiser headphones model HD414. The subjects' responses were recorded on tape for later analysis. The following are three examples of the Cryptic sentences.
Each sentence will be represented in the following way:

a) Swedish orthographic representation
b) English translation - word for word
c) broad Swedish phonetic transcription (Stockholm dialect)
d) Cryptic phonetic transcription

1.a) Inte längre ung men inte heller gammal.
   b) not anymore young but not either old.
   c) inte le`ngre en men inte hel`er gamal
   d) inte lengre ung men inte xeler gamal

2.a) Alla träd som hade skuggat kontorets brädväggar var nedsågade.
   b) all (the) trees which had shaded the office's walls were sawed down.
   c) ala trE:a som hEnE:e skuyat kantbradz baradvE:ar var
   d) aE: aE:ar som xE: E:e skuyot kEntorE:rdz bErE:vdE:ar var

3.a) Kontorsbyggnaden liknade barrackerna med platt tak och gula 
   b) The office building resembled the barracks with (a) flat roof
      and yellow partitions.
   c) kontu:zbyggnaE:n liknadE bak`arkE:na med plat tak ok gu:la
   d) kEntorzbE:gnadzE:en lE:gnadz bErE:karzE:ne med plat tak og gu:la

15 native speakers of Swedish were tested. The sentences can be 
grouped into three categories according to the type of responses they 
elicited: those which everyone could interpret correctly (e.g. sent. 1 
above); those which received no correct responses at all, not even 
single words (eg. sent. 3 above); those which were only partially 
correctly understood (eg. sent. 2 above). The majority of the
sentences fell into the last category.

The following are examples of incorrect interpretations by the subjects and possible explanations for these responses.

1. a) Det var en av dom som pratade med mamma om det.
   b) It was one of them who talked with Mom about it.
   c) de:t va:z en a:v dom som pru:ita: me:d mama om de:t
   d) de:d var en av dom som pørta a: med máma om de:t

   resp: Det var en av dom som börja(de)* tala med mamma om det.
   (it was one of them who began to talk with Mom about it.)

   The only word which is incorrectly interpreted in this response, is "pratade". This has instead become "börja(de) tala". A possible explanation is that the subject perceived the series of segments for "pratade" as: unaspirated labial stop + diffuse vowel + /r/ + diffuse vowel + /t/ + /a/ + voiced continuant consonant + diffuse vowel.

   This results in "börja(de) tala", which is in accordance with the number of syllables and stress pattern of the Cryptic pronunciation of "pratade":[pørta:ðː]. The phrase "började tala" also fits in the grammatical and semantic context surrounding it.

   In general, the responses of the subjects showed clearly the importance of semantic context in interpreting phonologically "distorted" sentences. Among the incorrect interpretations we find words that are not only phonetically similar to the actual ones, but

*The Swedish preterit ending (-de) is often deleted in casual speech when concatenated with a bisyllabic verb stem ending in a vowel.
also those which are semantically appropriate. An example of one word in a sentence influencing the interpretation of another word in the same sentence is as follows:

2.a) Alla träd som hade skuggat kontorets brädvåggar var nedsågade.

b) All the trees which had shaded the office's wooden walls were sawed down.

c) alla träd som hade skuggat kontorets brädvåggar var nedsågade.

d) alla träd som hade skuggat kontorets brädvåggar var nedsågade.

resp: Alla träd som ... skogen .......... [sku:gen] the forest

For many of the subjects, we found that if the listener perceives the beginning of the sentence [alla träd] correctly as "alla träd" (all (the) trees), he will tend to interpret [skuggat] as "skogen" (the forest) instead of "skuggat" (shaded). Observe that the concept of "forest" is within the same semantic field of "trees". By choosing "the forest", the listener's interpretation of the remaining parts of the sentence will be disturbed. Note that the Cryptic word [x:da] (Swe. "hade", Eng. "had") was not correctly interpreted here, though for several previous and following occurrences, it was. Therefore, the problem here cannot be one of sentence segmentation, i.e., that the listener did not know where the word began or finished, for as we see, the end and beginning of the preceding respective following word were correctly identified.

* Five "....." or more = no response, three "...") = remaining part of the sentence is unnecessary for the particular illustration
Such in comprehensions can be explained by the ungrammaticality of "hade" appearing between "som" and "skogen" which results in a refusal to believe that "hade" had really been said.

That the utilization of semantic context across sentences was an important strategy for interpreting the foreign accent is evident in many of the responses. Here we will see how the interpretation for the sentence in example 2 above has influenced the response to a sentence coming soon afterwards.

3.a) Ett halvår till skulle byggnaderna stå kvar.

b) A half year more would the buildings remain standing.

c) et hálvôːzi til skólè býgnadənə stoː kvaːz

d) et xálvar til skúle bęgnaděrə sto kavár

resp: .....................stockar (="logs")

Two sentences after the one in example 2, a segmentation problem has occurred. The listeners did not realize that the sequence [stókåvár] actually consisted of two words. This, coupled with the listeners' awareness of the content of the earlier sentence ("trees" and an incorrect response "forest") has led to the interpretation of "stå kvar" as the phonetically similar "stockar" [stôkåz] .

In the responses, we found that utilizing grammatical context was also a means for interpreting the broken Swedish. It appears that the listeners strive for grammatical correctness, which in turn governs the phonological interpretation.
Because of the listeners' incorrect interpretation of [xans skúlar] as [han skéle], the demand for grammatical correctness forces him to ignore the preterit ending in the Cryptic variant of "hade" - [xáða]. "hade" therefore becomes "ha", in spite of the fact that the word [xáða] was interpreted correctly prior to as well as after this particular occurrence.

CONCLUSION

This study, being a pilot study, is not by itself adequate to provide definitive answers to the questions which we posed in the introduction. We have, for instance, not been able to establish a precise hierarchy of difficulty for interpreting the different pronunciation errors. Our impression is that the errors which are most disturbing for the listener are change of word stress and change of the number of syllables in a word by vowel insertion. Sentences containing several instances of these two types of pronunciation errors proved to be the most unintelligible for all of the subjects. It appears that there is a tendency for the listeners to retain the prosodic pattern of the Cryptic word and attempt to match a Swedish word (or words) with similar segmental composition to this pattern. Why the suprasegmental structure of a Cryptic word is more resistant
to reinterpretation than the segmental is beyond the scope of the present study.

Another observation which we have made is that there is often a "domino effect" of unintelligibility. That is, if a particular word in a sentence is either unintelligible or interpreted incorrectly, it will affect the perception of the words immediately following. The interference by one word on the identification of the successive words occurs even if the latter are words which the subjects ordinarily has not had difficulty comprehending. If the word is unintelligible, then the problem is one of segmentation, i.e., not knowing where one word ends and the next begins. If a word is misinterpreted, and if the interpretation is not grammatically and semantically appropriate to the remaining text, the listener will be thrown off the track for the words which follow (and in some cases precede) it.

A string of unintelligible words may cause a "shock effect" which can linger on for several sentences. The listener becomes distressed because he cannot understand anything of what he hears, causing him to give up. Not until a very easily understood sentence is heard will the subject start responding again.

The questions of whether a listener will form phonological rules for interpreting the pronunciation errors and how long it takes to do so for each particular error type, are not possible to answer based on this pilot study alone. To do so, we have continued the investigation by designing and conducting a new test, based on the observations made from the study just presented. In the new test, we have controlled the occurrence of the different error types in the following way:

The Cryptic phonological rules are grouped into three main categories:

1. Segmental rules (specifies segments, e.g. /h/ → [x])
2. Prosodic rules (stress is always placed on the penultimate)
3. Vowel insertion (change of syllable structure by insertion of a neutral vowel)

By applying rules from only one or a combination of two or three categories onto separate sentences, we can better study the degree of difficulty caused by the individual categories. We have also lengthened the test from 30 sentences to 48 sentences to increase the exposure of the subjects to the Cryptic accent. Our hope is that with this increased exposure, a learning effect will be easier to discern. To aid in the establishing of a possible learning effect, we have added a list of single words to the end of the task. By allowing a control group to listen to this list alone, without a Cryptic text preceding it, we can compare the list's intelligibility for the two groups. If the group which listens to the text before the word list performs better than the other group, we can more precisely state whether a learning of the Cryptic phonologic system has occurred, and what the extent of this learning has been.
THE USE OF SYNTHETIC SPEECH IN THE CORRECTION OF DEAF SPEECH

Anne-Marie Ö ster

Abstract

The speech of the congenitally deaf is characterized by many errors that affect speech intelligibility. It is likely, however, that these errors do not occur randomly but follow specific rules that are speaker dependent.

At the Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, a text-to-speech system has been developed. The system consists of a set of computer rules that transforms the written text into synthetic speech by means of phonetic rules. With this technique it is possible to study the effect of a specific type of phonetic error on intelligibility.

Speech samples from three deaf children were individually analyzed for errors in vowels, consonants and prosody. Based on this analysis, a phonetic system for each child was established. A group of normal-hearing subjects listened to the synthetic deaf speech and were asked to write down all of the words that they could understand. The goal was to classify the speech errors of each child in some order of precedence according to intelligibility.

The results of this test can be used to establish the relative impact on intelligibility of different types of speech errors and to develop an individualized program for speech improvement.
Inledning


Syftet med detta arbete var att undersöka olika talfelns relativa inflytande på talförståeligheten hos tre barndomsdöva barn. Det viktigaste var alltså att försöka ta reda på de talljud vars uttal är viktigast för förståeligheten, dvs vilka talfel som bör korrigeras först, eftersom de försämrar talet mest. Att rangordna talfelen efter avtagande förståelighet kan förhoppningsvis vara till nytta för pedagogen och ge vissa rekommendationer för i vilken ordningsföljd talkorrektionen bör utföras. För att uppnå en effektiv talkorrektion och talträning är det alltså viktigt att utarbeta metoder, som utvärderar barnens tal och på ett enkelt sätt summerar deras talfel och metoder som bestämmer dessa talfels inverkan på förståeligheten.

regler konstrueras för önskade uttalsförändringar.

Vi valde att använda regelsyntes eftersom vi ansåg att våra försökspersoner hade väletablerade talvänor dvs att ett talljud i liknande fonetisk och lingvistisk kontext realiserades på samma sätt. Döva talare lär in talljud med hjälp av uttalsregler och använder dessa regler konsekvent i sitt tal. Den använda syntestekniken, som baseras på fonologiska regler, fann vi vara ett lämpligt verktyg trots brister i naturligheten.

Att arbeta med annan typ av talsyntes t.ex. LPC-syntes, finsyntes eller ihopklippt tal ansåg vi olämpligt av följande anledningar:

LPC-syntes eller talkodning hade i första hand möjliggjort ändringar i den suprasegmentella produktionen realiserade genom variationer i tempo och grundton. Konsonant- och vokalkvalitetsändringar i den segmentella produktionen skulle varit svåra att generera.

Finsyntes skulle för denna undersökning ha krävt en alltför stor arbetsinsats. Denna teknik saknar lingvistiskt sammanhang och kan inte beskriva barnens tal genom generaliseringar utan endast genom ett fåtal utsagor.

Ihopklippt tal ger dålig prosodi och osammanhängande talkvalitet med bristande koartikulationseffekter.


figur 1.
Resultatet från denna undersökning visade att:

1) ett tal som innehåller enstaka prosodiska fel eller ett ensamt vokal- eller konsonantfel skulle ge ett förståeligt tal.

2) Då vokal- eller konsonantfel kombineras med varandra eller då talet störs av mellanvokaler (störande vokalliknande ljud som uppstår på grund av de långsamma artikulatoriska rörelserna) sjunker förståeligheten till svår förståeligt.

För att ytterligare studera olika talfels relativa inflytande på förståeligheten beslöt vi att genomföra en fallstudie av tre gravt hörselskadade barns tal.

Talstatus av tre gravt hörselskadade barns tal


![Figur 3.](image)

Vi spelade in barnens tal på band när de läste en- och flersstäviga ord och en sammanhängande text. Orden representerade alla svenska vokaler, klusiler, frikativor, nasaler, likvider, supradentaler och konsonantförbindelser i initial, medial och final ställning. Den sammanhängande texten analyserades för att bestämma barnens grundtonsfrekvens och dess variation, pausering, taltempo och betoning (se bilaga 1 för material för talstatus). Före inspelningen av varje lista övningsläste barnen och svåraste ord förklarades.

En fonetiskt tränad lyssnare transkriberade de tre barnens segmentella produktion, dvs vokaler och konsonanternas uttal och sammanställde barnens uttalsfel. Den suprasegmentella produktionen, dvs tidsrelationer och grundtonsvariationer, analyserades med hjälp av mingogram. I stället för att tala om barnens prosodi föredrar vi att tala om den suprasegmentella produktionen, eftersom dövas tal på många sätt skiljer sig kraftigt från normalhörandes tal. Dövas tal karaktäriseras i hög grad av felaktiga tids- och längdförhållanden t ex en genomgående segmentförlängning, pausering mellan ord eller stavelser,
störande övergångsljud på grund av artikulatorernas långsamma rörelser, brist på koartikulation och långa ocklusioner (tillslutningen av luftströmmen i talröret innan munhålan öppnas vid produktionen av klusilerna p, t och k) vilka samtliga bidrar till ett långsamt och monotont tal. Nedan följer en systematisk beskrivning av de tre barnens talfel. Talfelen låg sedan till grund för de fonetiska regler som konstruerades för att styra talsynthesapparatens artikulation.

Kartläggning av barnens talfel

VOKALER

Döva talares vokalproduktion är ofta kraftigt reducerad (se Martony, Nordström & Öster: Dövas Vokalproduktion, 1975). De behärskar avläsbbara egenskaper såsom läpprundning och käköppning tämligen säkert men har svårigheter att producera det som inte är avläsbart såsom tungställning. De förväxlar inte bara vokaler som är artikulatoriskt lika, t ex I och E, som båda bildas genom att tungan förs långt fram och högt upp mot gommen, utan även sådana som är artikulatoriskt olika, framför allt rundade vokaler.

Figur 4 visar de tre barnens vokaler placerade i två vokalfyrsidingar, som visar orundade respektive rundade svenska vokaler. De inringade tecknen i figuren motsvarar de vokalljud som barnen behärskar. Övriga tecken motsvarar de svenska vokalernas tunglägen.


Vokalfyrsidingen är följaktligen ett bra pedagogiskt hjälpmedel vid talkorrektion för att beskriva tungans roll vid vokalartikulation. Tabell 1, 2 och 3 ger en utförligare beskrivning av de fonetiska tecken som använts i figur 4 samt redovisar hur barnen reducerar uttalelsen av de svenska vokalerna.
Figur 4.
Exempel | IPA | Barn A:s representation
---|---|---
SIL | i: | ĕ
SILL | e: | ē
VET | e: | ē
SÄL | e: | ē
HETT | | ē
HÄR | æː | æː
HERR | æː | æː
SYL | yː | òː
HUS | æː | æː
FÖL | øː | æː
FÖR | æː | æː
FÖRR | | æː
SYLL | æː | æː
HUND | æː | æː
ROTT | uː | ûː
GÅTT | æː | ãː
GÅ | æː | ãː
HATT | aː | ãː
HAT | aː | ãː

Tabell 1. Barn A:s vokalsystem.

**Barn A:** reducerar de svenska vokalerna till sex korta och två långa vokaler. Samtliga vokaler är nasalade. Alla slutsna vokaler uttalas halvöppet.

<table>
<thead>
<tr>
<th>Exempel</th>
<th>IPA</th>
<th>Barn B:s representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIL</td>
<td>i:</td>
<td>ə</td>
</tr>
<tr>
<td>SILL</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>VET</td>
<td>e:</td>
<td>e:</td>
</tr>
<tr>
<td>HETT</td>
<td>ε</td>
<td>ε</td>
</tr>
<tr>
<td>SÅL</td>
<td>ε:</td>
<td>ε:</td>
</tr>
<tr>
<td>HERR</td>
<td>aε</td>
<td>aε:</td>
</tr>
<tr>
<td>HÄR</td>
<td>aε:</td>
<td></td>
</tr>
<tr>
<td>SYL</td>
<td>y:</td>
<td>y(ː)</td>
</tr>
<tr>
<td>SYLL</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>FÖL</td>
<td>ø:</td>
<td>ø:</td>
</tr>
<tr>
<td>FÖR</td>
<td>ø:</td>
<td>ø:</td>
</tr>
<tr>
<td>FÖRR</td>
<td>øe</td>
<td>øe:</td>
</tr>
<tr>
<td>HUS</td>
<td>u:</td>
<td>u(ː)</td>
</tr>
<tr>
<td>ROTT</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>ROT</td>
<td>u:</td>
<td>θ(ː)</td>
</tr>
<tr>
<td>GÅ</td>
<td>o:</td>
<td></td>
</tr>
<tr>
<td>GÅTT</td>
<td>j</td>
<td></td>
</tr>
<tr>
<td>HUND</td>
<td>θ</td>
<td></td>
</tr>
<tr>
<td>HAT</td>
<td>a:</td>
<td>a:</td>
</tr>
<tr>
<td>HATT</td>
<td>a</td>
<td>a:</td>
</tr>
<tr>
<td>Exempel</td>
<td>IPA</td>
<td>Barn C:s representation</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>--------------------------</td>
</tr>
<tr>
<td>SIL</td>
<td>iː</td>
<td>iː</td>
</tr>
<tr>
<td>SYL</td>
<td>yː</td>
<td>yː</td>
</tr>
<tr>
<td>SILL</td>
<td>iː</td>
<td>iː</td>
</tr>
<tr>
<td>SYLL</td>
<td>yː</td>
<td>yː</td>
</tr>
<tr>
<td>HUND</td>
<td>θ</td>
<td>θ</td>
</tr>
<tr>
<td>VET</td>
<td>eː</td>
<td>eː</td>
</tr>
<tr>
<td>HETT</td>
<td>e</td>
<td>e</td>
</tr>
<tr>
<td>SÅL</td>
<td>eː</td>
<td>eː</td>
</tr>
<tr>
<td>HÄR</td>
<td>æː</td>
<td>æː</td>
</tr>
<tr>
<td>HERR</td>
<td>æ</td>
<td>æ</td>
</tr>
<tr>
<td>FÖL</td>
<td>øː</td>
<td>øː</td>
</tr>
<tr>
<td>FÖRR</td>
<td>æː</td>
<td>æː</td>
</tr>
<tr>
<td>FÖR</td>
<td>æː</td>
<td>æː</td>
</tr>
<tr>
<td>ROTT</td>
<td>u</td>
<td>uː</td>
</tr>
<tr>
<td>GÅTT</td>
<td>øː</td>
<td>øː</td>
</tr>
<tr>
<td>HUS</td>
<td>uː</td>
<td>uː</td>
</tr>
<tr>
<td>ROT</td>
<td>uː</td>
<td>auː</td>
</tr>
<tr>
<td>GÅ</td>
<td>õː</td>
<td>õː</td>
</tr>
<tr>
<td>HAT</td>
<td>åː</td>
<td>åː</td>
</tr>
<tr>
<td>HATT</td>
<td>å</td>
<td>å</td>
</tr>
</tbody>
</table>

Tabell 3. Barn C:s vokalsystem.

Barn C: har så många som tolv långa vokaler och endast en kort vokal. Orundade vokaler nasaleras och vissa rundade vokaler diftongeras.
KONSONANTER


Vid övergång mellan nasalkonsonant och icke nasal vokal är det vanligt att en extra konsonant uppstår med samma artikulationsställe som nasalen på grund av alltför långsam velumrörelse. Nasalen /m/ t ex uttalas då istället som /mb/.

Konsonanter: Barn A

Konsonanter: Barn B

## Konsonanter: Barn B

### Klusiler:

<table>
<thead>
<tr>
<th>Konsonant</th>
<th>Initial</th>
<th>Medial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>_theme</td>
<td>p, b</td>
<td>[p]</td>
<td>[b]</td>
</tr>
<tr>
<td>_theme</td>
<td>t, d</td>
<td>[t]</td>
<td>[d]</td>
</tr>
<tr>
<td>_theme</td>
<td>k</td>
<td>[k]</td>
<td>[k]</td>
</tr>
<tr>
<td>_theme</td>
<td>g</td>
<td>[g]</td>
<td>[g]</td>
</tr>
</tbody>
</table>

### Frikativor:

<table>
<thead>
<tr>
<th>Konsonant</th>
<th>Initial</th>
<th>Medial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>_theme</td>
<td>f, v</td>
<td>[f]</td>
<td>[v]</td>
</tr>
<tr>
<td>_theme</td>
<td>s, sj</td>
<td>[s]</td>
<td>[sj]</td>
</tr>
<tr>
<td>_theme</td>
<td>tj</td>
<td>[tj]</td>
<td>[tj]</td>
</tr>
<tr>
<td>_theme</td>
<td>j</td>
<td>[g]</td>
<td>[g]</td>
</tr>
<tr>
<td>_theme</td>
<td>h</td>
<td>[h]</td>
<td>[h]</td>
</tr>
</tbody>
</table>

### Likvider:

<table>
<thead>
<tr>
<th>Konsonant</th>
<th>Initial</th>
<th>Medial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>_theme</td>
<td>l</td>
<td>[l]</td>
<td>[l]</td>
</tr>
<tr>
<td>_theme</td>
<td>r</td>
<td>[r]</td>
<td>[s]</td>
</tr>
</tbody>
</table>

### Nasaler:

<table>
<thead>
<tr>
<th>Konsonant</th>
<th>Initial</th>
<th>Medial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>_theme</td>
<td>m</td>
<td>[m]</td>
<td>[m]</td>
</tr>
<tr>
<td>_theme</td>
<td>n, ng</td>
<td>[g]</td>
<td>[g]</td>
</tr>
</tbody>
</table>

Konsontantförbindelser med retroflex artikulation (bäckåtbojda tungspets som i förs, fort, bord, pårla, hörn):

- **rs** → [s[k]
- **rt** → [t[k]
- **rd** → [d[k]
- **rl** → [l[g]
- **rn** → [n[g]

Barn B uttalar klusilerna som tonande i initial ställning. Om klusilerna förekommer sist i ett ord uttalar de tonlöst utom /g/ som i stället nasaleras. Dessutom aspireras klusilerna kraftigt, när de förekommer sist i ett ord. s-, sj- och tj-ljud är mycket otydliga och uttalar som sj-ljud föregåtget eller efterföljt av /k/. Många av konsonantljuden bildas långt bak i munnen. Uttalet av /r/ påminner om det tyska eller skånska uttalet. De retroflexa konsonantförbindelserna uttalar han liksom barn A var för sig.
**Konsonanter: Barn C**

**Klusiler:**

<table>
<thead>
<tr>
<th>Konsonanter</th>
<th>Start</th>
<th>Medel</th>
<th>Slutfase</th>
</tr>
</thead>
<tbody>
<tr>
<td>p, b</td>
<td>[p]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t, d</td>
<td>[t]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k, g</td>
<td>[k]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Frikativor:**

<table>
<thead>
<tr>
<th>Konsonanter</th>
<th>Start</th>
<th>Medel</th>
<th>Slutfase</th>
</tr>
</thead>
<tbody>
<tr>
<td>f, v</td>
<td>[f]</td>
<td>[f]</td>
<td>[f]</td>
</tr>
<tr>
<td>s</td>
<td>[s]</td>
<td>[s]</td>
<td>[s]</td>
</tr>
<tr>
<td>sj, tj, j</td>
<td>[sj]</td>
<td>[sj]</td>
<td>[sj]</td>
</tr>
<tr>
<td>h</td>
<td>[h]</td>
<td>[h]</td>
<td>[h]</td>
</tr>
</tbody>
</table>

**Nasaler:**

<table>
<thead>
<tr>
<th>Konsonanter</th>
<th>Start</th>
<th>Medel</th>
<th>Slutfase</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>[mb]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>[nd]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ng</td>
<td>[ng]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Konsonantförbindelser med retroflex artikulation (bakåtböjd tungspets som i fär, fort, bord, pär[a], hörn):

<table>
<thead>
<tr>
<th>Likvider</th>
<th>Start</th>
<th>Medel</th>
<th>Slutfase</th>
</tr>
</thead>
<tbody>
<tr>
<td>l</td>
<td>[l]</td>
<td>[l]</td>
<td>[l]</td>
</tr>
<tr>
<td>r</td>
<td>[r]</td>
<td>[r]</td>
<td>[r]</td>
</tr>
<tr>
<td>nl</td>
<td>[nl]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SUPRASEGMENTELL PRODUKTION (tidsrelationer och $f_0$)

I figur 5 visas typiska skillnader mellan normaltal och dövtal. Den heldragna linjen visar grundtonsfrekvensens variation (satsintonationen) uppmätt genom mingogram. Uttrycket är "har du ingen näsduk alls?". Den prickade linjen återger tonlösa ljud.

Den låga talastigheten, pauseringen mellan varje stavelse och frånvaron av satsintonation är typiskt för den döve talaren. Den höranande talaren betonar orden "ingen" och "alls" och sammanbinder sitt tal till en enhet. Om ord behandlas som separata enheter blir betoningens monoton och talet sönderhackat. En sådan talteknik blir mycket tröttande för såväl talare som lyssnare.


Figur 5.
Suprasegmentell produktion: Barn A


BARN A
Pausering mellan varje stavelse
Hög grundton
Grundtonssänkning i slutet av varje vokal
Förlängd segmentlängd

Suprasegmentell produktion: Barn B


BARN B
Pausering efter två ord
Förlängning av första - sista stavelsen
Vokalberoende grundton:
Höjning vid Ö Sänkning vid E
Ö Ä
U: A
I Å
O U
O:
Suprasegmentell produktion: Barn A


BARN A
Pausering mellan varje stavelse
Hög grundton
Grundtonssänkning i slutet av varje vokal
Förlängd segmentlängd

Suprasegmentell produktion: Barn B


BARN B
Pausering efter två ord
Förlängning av första – sista stavelsen
Vokalberoende grundton:
Höjning vid Y Sänkning vid E
Ö Å
U: A
Å Å
O U
O:
Suprasegmentell produktion: Barn C

Detta barn har normalt röstläge. Taltempot är påfallande långsamt och han gör uppehåll efter varje ord och avslutar varje fras med en kraftig förlängning av sista stavelsen. Dessutom förlänger han ocklusionsfasen (tillslutningen av luftpassagen i talröret före "explosionen" då munnen öppnas) vid produktion av klusilerna /p, t, k/.

BARN C
Pausering mellan varje ord
Sista stavelsen extremt lång
Långa ocklusioner
Långsam taltempo
Lyssningstest

För varje barns tal konstruerades tre starkt förenklade uttalsbeskrivningar som illustrerade barnens talfel. Dessa tre beskrivningar består av fonetiska datorregler som styr talsyntesapparaten för talsynthes. Varje sådan regel översätter respektive barns uttal från skrift till tal. Följaktligen var de tre barnens uttalslexika samman­satta av vokal-, konsonant- och suprasegmentella regler grundade på det för varje barn genomgångna talstatusmaterialet.

Material

Testmaterialet bestod av 96 meningar (se bilaga 2). Varje mening innehöll fyra nyckelord. I en typisk mening som "ficklampa lyste svagt i mörkret" är de underströkna orden nyckelord. Prepositioner, artiklar, adverb och tempus betraktades som redundanta, dvs om försökspersonen uppfattade "ficklampa lysa svagt mörker" kan han med hjälp av sin språkliga kompetens bygga upp den rätta meningsstrukturen. Lyssnaren utnyttjar språkets redundans för att fylla i det som saknas. Vid analysen av lyssningstestet räknades därför endast varje korrekt uppfattat nyckelord.

Av de 96 testmeningarna återger meningarna 1-32 barn A:s tal, 33-64 barn B:s och 65-96 barn C:s tal. Varje talfel och kombinationer av talfel presenterades i grupper om fyra meningar. Talfelen grupperades på följande sätt:

0 - utan fel
Vokal fel
Konsonant fel
Suprasegmentella fel
Vokal fel + Suprasegmentella fel
Konsonant fel + Suprasegmentella fel
Vokal fel + Konsonant fel
Vokal fel + Konsonant fel + Suprasegmentella fel

\[ X \times 4 = 32 \]


Försökspersoner

Försökspersoner var 21 studerande vid inst. för lingvistik vid Stockholms Universitet. Lyssningstestet utfördes vid språklaboratoriet vid ett gemensamt tillfälle. Försökspersonerna lyssnade via
hörlurar och antecknade svaren på ett testformulär. Testet inleddes med åtta övningsmeningar utan fel för att de skulle vänja sig vid det syntetiska talet.

**Resultat**

Figurerna 6, 7 och 8 visar resultat av lyssningstestet. För varje barn är talfelen ordnade i avtagande förståelighet. Resultaten är indelade i tre grupper efter grad av förståelighet. Vid 100-50% rätt uppfattade nyckelord bedöms talet som förståeligt. De talfel som ligger inom 50-25% förståelighet påverkar talet så att det blir svår­förståeligt och de som faller inom gruppen 25-0% gör talet oförståe­ligt.

De enstaka talfelen hos Barn A (figur 6) inverkar på förståeligheten i högre grad än hos Barn B. När de segmentella felen (vokaler och konsonanter) förekommer samtidigt, blir talet svår­förståeligt till o­förståeligt och vid ytterligare tillägg av respektive suprasegmentella fel, blir båda barnens tal fullständigt oförståeligt.

Figur 8 visar att Barn C:s tal är helt förståeligt. För de fyra meningar som innehöll hans samtliga talfel (vokal-, konsonant- och suprasegmentella fel) var förståeligheten 64%.
Figur 7.

Figur 8.

<table>
<thead>
<tr>
<th>Vokalfelens effekt</th>
<th>Konsonantfelens effekt</th>
<th>Suprasegmentella felens effekt</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BARN A</strong></td>
<td><strong>TALFEL MED</strong></td>
<td><strong>TALFEL MED</strong></td>
</tr>
<tr>
<td><strong>OKALFEL</strong></td>
<td><strong>VOKALFEL</strong></td>
<td><strong>VOKALFEL</strong></td>
</tr>
<tr>
<td>V 71%</td>
<td>V 71%</td>
<td>V 71%</td>
</tr>
<tr>
<td>VK 12</td>
<td>KS 51</td>
<td>KS 51</td>
</tr>
<tr>
<td>VS 18</td>
<td>VS 18</td>
<td>VS 18</td>
</tr>
<tr>
<td>VKS 17</td>
<td>VKS 17</td>
<td>VKS 17</td>
</tr>
<tr>
<td><strong>TALFEL UTAN</strong></td>
<td><strong>VOKALFEL</strong></td>
<td><strong>VOKALFEL</strong></td>
</tr>
<tr>
<td>G 97%</td>
<td>K 38%</td>
<td>K 38%</td>
</tr>
<tr>
<td>S 66%</td>
<td>S 66%</td>
<td>S 66%</td>
</tr>
<tr>
<td>VS 18</td>
<td>VS 18</td>
<td>VS 18</td>
</tr>
<tr>
<td>KS 51</td>
<td>KS 51</td>
<td>KS 51</td>
</tr>
<tr>
<td>VK 12</td>
<td>VK 12</td>
<td>VK 12</td>
</tr>
</tbody>
</table>

**Figur 9.**

Vid enbart vokalfel är talet fortfarande förståeligt (71%). Vid tillskott av andra samtidiga fel sker en kraftig försämring av förståeligheten. I samtliga fall blir talet oförståeligt. Enbart konsonantfel påverkar talet i hög grad. Förståeligheten sjunker från 97% till 38%. En kombination av detta barns vokal- och konsonantfel är allvarlig och talet blir helt oförståeligt (12%). Barnets supra-
segmentella fel påverkar inte talet i lika hög grad men i kombination med vokalfel får det dock allvarliga följder. Förståeligheten sjunker då från 71% för enbart vokalfel till endast 18%. I övriga fall ökar förståeligheten vid tillägg av suprasegmentella fel tillsammans med konsonantfel och med konsonant- och vokalfel. Detta kan möjligen bero på den förlängda segmentlängden, som ingår i det suprasegmentella felet. Härigenom blir talet långsammare, vilket kanske i dessa fall innebär att lyssnaren hinner segmentera och tolka innehållet bättre.

**Förslag till korrektion av Barn A:s tal**

Figur 10. visar på samma sätt vokal-, konsonant- och de suprasegmentella felens effekt på Barn B:s talförståelighet.

**Förslag till korrektion av Barn B:s tal**


Förslag till korrektion av Barn C:s tal:

Först bör konsonantfelet korrigeras. Förståeligheten ökar till 91% utan konsonantfel (VS). Därefter bör vokalerna tränas, vilket skulle innebära att förståeligheten ökar till 96% (S). Även i detta fall skulle träningen av den suprasegmentella produktionen komma i sista hand.

Avslutning

För att talundervisningen skall leda till ett förståeligt tal måste talträningen vara effektiv. Ett effektivt talträningsprogram kräver, som tidigare påpekats, att man dels har metoder för att kartlägga elevens uttalsfel, dels metoder för att bedöma dessa fels relativa betydelse för förståeligheten. Detta arbete har varit ett försök att i tre fall bedöma talet hos tre gravt hörselskadade barn från Manillaskolan samt ge rekommendationer till individuell korrektion av deras tal genom att visa på de fel, som försämrar förståeligheten mest, och som därför bör korrigeras först.
Litteratur


BILAGA 1. Material för talstatus

VOKALER:

1. APA  2. KATT  3. KO
4. BLOMMA  5. GUL  6. TUPP
7. LÅS  8. BOLL  9. EK
10. EKORRE  11. BIL  12. RING
13. Fyr  14. NYCKEL  15. SKÅR
16. HÄST  17. RÖD  18. ÖRA

KLUSILER:

1. PIPA  2. LÄPPAR  3. KOPP
4. BUR  5. GUBBE  6. NÄBB
7. TÄG  8. MATT A  9. BÅT
10. DOCKA  11. LÅDA  12. VED
13. KAM  14. KAKA  15. BOK
16. GAPA  17. ÖGA  18. ÄGG

NASALER:

1. MOROT  2. LAMPA  3. LAMM
4. NÅL  5. PENNA  6. INDIAN
7. GUNGA  8. SÄNG  9. BALLONG
FRIKATIVOR:
1. _SOL_ 2. _SIL_ 3. _MÖSSA_
4. _BUSS_ 5. _POLIS_ 6. _FOT_
7. _SOPPA_ 8. _GIRAFF_ 9. _VAS_
10. _LUVA_ 11. _HOV_ 12. _SKJORتا_
16. _DUSCH_ 17. _KÄLKE_ 18. _KJOL_
22. _VARG_ 23. _HATT_ 24. _OHYRA_

TREMULANT - LATERAL:
1. _RÄTТА_ 2. _PÄRON_ 3. _DÖRR_
4. _Får_ 5. _LOK_ 6. _BULLE_
7. _PALL_

KONSONANTFÖRBINDELSER MED RETROFLEX ARTIKULATION:
1. _KORT_ 2. _BORD_ 3. _GARN_
4. _FORS_ 5. _FÄRLA_

KONSONANTFÖRBINDELSELAR:
1. _PLÄSTER_ 2. _TABLETTER_ 3. _FLYGPLAN_
4. _TVÅL_ 5. _KLOCKA_ 6. _KNAPP_
7. _KVITTO_ 8. _SKO_ 9. _FISK_
10. _SPIK_ 11. _SVAMP_ 12. _STOL_
13. _OST_ 14. _SMÖRGÅS_ 15. _SNÖ_
"LOTTA VAR SNUVIG OCH FICK INTE FÖLJA MED MAMMA OCH HANDLA. DÅ BLIR JAG ARG! JAG VILL GÅ TILL AFFÄREN, SA LOTTA. I AFFÄREN SNORADE OCH SNORADE HON, OCH TILL SIST FRÅGADE EN TANT: HAR DU INGEN NÄSDUK ALLS? JO, MEN DEN VILL JAG INTE LÅNA UT TILL DIG, FÖR JAG KÄNNER DIG INTE, SVARADE LOTTA."
BILAGA 2. Meningar för lyssningstestet

TESTMENINGAR FÖR BARN A:S TALFEL:

M"AM`AN T`O:G SIN+ D"ÄT`Ä4R I:+ "Ö3R`AT
D"Ö4R#KL`ÅKAN HA:R+ EON+ J`ÄL SINGN`Ä:L
K`UNDEN P"Ä:R+ AT+ FR`UKTEN Ä:+ R"UT`EON
DEON+ G"AML`Å L`UP`AREN B`OD`E I:+ H`U:SET
FR`ISK L`UFT, Ä:+ M"ATSJ`Ö:N JÖ3R+ N"YT`Ä
J`U:DET TRÄNGÄ4R "INT`E FR`AM
B`ÄKEN I:+ D`Å:LEN SV"ÄM`ADE Ö:VEOR
SM`Ä: B`Å:2N SITÄ4R I:+ KN`Ä:T
V`Ä:GEN TIL+ ST"ÄK#H`ÄLM Ä:+ B`AK`IG, Ä:+ KR`O`K`IG
PAR`I:S BR`O:AR L"ÄK`AR TUR`ISTÄ4R
PR`I:S`E2NA ST`I:GÄ4R FÖ3R+ V`ÄR`J`E0 `Ä:R
`Ö:RMEN R"INGL`ADE I:+ DE+ H"Ö:G`A GR`Ä:SET
DE+ N"Y`A KL`ISTRET F`ÄSTÄ4R P"Ä:+ `ALT
"E`V`A HA:R+ TAPAT N`YK`ELN TIL+ S`YKELN
F`I:O:L, Ä:+ TR`UMP`E:T Ä:+ G`AML`Å INSTR"UM`ENT
L"ÄP`OR Ä:+ G`ÅNSK`Å+ S`ÄLS`YNTA J`U:R
M`ASKROS`02NA L`Y:SÄ4R G`U:LT I:+ S`O:LEON
"ASK`AN E`K`ADE MELAN B`Ä4RJEN
`Y:STAD Ä:+ SV`Ä4RJES S`Y:DL`IGASTE ST`A:D
B`I:LEN FIK+ EON+ B"UKL`Å P`Ä:+ SJ`Ä4RMEN
GR`UPEN BEST`ÄR A:V+ DR`A:G#SP`E:L, Ä:+ J`IT`AR
K`ALV`02NA L`Y:SÄ4R M`UNTÄ42T I:+ S`O:LEON
P`ÄJKEN BLÖ:DÄ4R "ÄFT`A N`Ä:S#BL`O:D
H`UNDEN SJ`ÄLÄ4R PÅ:+ GR`AN`ENS B`Å:2N
P`ÄJKA2NA ÄKTE SKR`ISK`Ö:R PÅ:+ `I:SEN
T`Å:GET KÄM+ EON+ T`IM`E FÖ3R+ T`I:D`IT
L`ÄNÄN Ä:+ EUR`O:PAR ST`Ö42S2T`Å ST`A:D
K`ÄTEN HA:R+ JÖMT SI:NA "UNG`AR FÖ3R+ R`Ä:VEN
"ASK`AN HA:R+ SLÅ:GIT N`E:R I:+ `E:KEN
TESTMENINGAR FÖR BARN B:S TALFEL:

K"AJS`A GR`Ä:T NÄ3R+ BAL`ÄNGEN SPR`AK
DEON+ KL"A:R`A R`ÖSTEN, H"Ö32D`ES L`ÄNGT
`ÄGEN R"AML`ADE I:+= G"A:T`AN, MEON+ H`ÖL
K`UNDE BET`A:LAR REKL`A:MN FÖ3R+ V"A:R`AN
J`Ö:2DEN Ä:+ NÄSTAN+ R`UND, SÄM+ EON+ B`ÅL
KAMP`A`JEN "A:V"SL´U:TAS R"E:D`AN PÅ:+ S`ÖNDA
K"A:R`IN "ARB`E:TAR PÅ:+ "AVD`E:LNING T`ÄLV
H`UNGÄ42N SL´E:T I:+= M"A:G`EON PÅ:+ V`ARJEON
P"EN`A, Ä:+ P"AP`EOR FINS PÅ:+ B´O:2DET
S`ÅKEOR Ä:+ "INT`E N"YT`IT FÖ3R+ T`ÄNDÄ42NA
ST´I:GEN SL"INGR`ÄDE SEJ+ UNDEOR+ TR´Ä:DEN
K"AL`E L"Ä:ST´E SI:NA L"ÄKS`OR SL"ARV´IT
TJ"Ö32SÅB`Ä3REN H"ÄNGD´E I:+ ST"O:R`A KL"A:S`AR
ALÅ+ TE`A:TRA2NA ST´ÄNGÄ4R PÅ:+ S"ÄM`AREN
B"A:G`ARENS B`Ö4R Ä:TÄ4R J"Ä32`NA B"UL`AR
ST"E:N`A2NA R"UL`ÄDE NE:RFÖ3R B`AK`EON
MUSIK`ANTEN SP"E:L`ÄDE EON+ GAMAL P"ÅLK`A
SKR´ISKÖ2NA BEHÖ:VÄ4R SL"I:P`AS "ÄFT`A
ST´ÄRKEN Ä:+ EON+ S"ALS`YNT F`Ä:GEL I:+ SK"Ä:N´E
R´ENGETN F´ÖL I:+ H"ÄFT`IGA SK`U:R`AR
GL´ASEN SMÄLTÅ4R SN´ABT I:+ S´Ø:LEON
SP"O:K´EN F´INS PÅ:+ GAMLA SL´ÅT
S`ÄF`AN Ä:+ BÅ:DE BEKV´ÄM, Ä:+ B´IL`IG
L"IL´AN HA`R T"AP´AT SIN+ F´Ö42S2T´A T´ÄND
M¨Ä:L´AREN HA`R STA:VAT F´E:L PÅ:+ SJ´YLSEN
V`A:SEN R`AML`ADE I:+ G´ÅLVET, Ä:+ SPR`AK
L"EJ´ÅNET R´Ö:T SÅ:+ B´U:REN SK`A:K`ADE
S`YKL´ING, Ä:+ T`EN´IS JE:R+ G´O:D M´ÅTSJ´O:N
ÄRK`EST`Ä42N SP¨E:L`ÄDE S`ÅRL´IG MU:SI:K
B´Å:TEM R`UL`ÄDE KR`AFT´IT I:+ ST´ÄRKEN
D´ÄFTEN A:V+ K`AF`E SPR`E:D SEJ+ I:+ H´U:SET

145
TESTMENINGAR FÖR BARN C:S TALFEL.

KL"AK'AN R'INGÄ4R NÄ3R+ R'ASTEON Ä+: SL'U:T
G"AML'EOR V"I:S'AS I:+ T"E:V'E:
F"Ä:GELN KÄM+ ME+ M'Ä:T Â+:T+ "UNG'Ä2NA
FL"IK'AN S'ITÄ4R, Â+: L'Ä:SÄ4R EON+ S"A:G'OB:O
T"I:O H"U:V'UDEN S"YNT'ES I:+ D'Ö4REN
P'ÄJKEN VA:R+ R'ÄD FÖ3R+ DEON+ SV"A2T'A TJ'U:REN
GR"ÄSP'ARVEN Ä+: EON+ L"I:T'EON F"Ä:GEL
POT'Ä:TIS SKA:+ K"O:K'AS, INAN+ DEON+ "Ä:T'ES
"ÄPL'ET Ä+: M"O:G'ET, Ä+: "ALD'ELER GR'Ö:NT
BJ'Ö4RKEN HA:R+ T"AP'AT ALA+ BL'Ä:DEN
B'ÄKEN M"YN'AR I:+ DEON+ GR'UND'Ä V'I:KEN
B'USEN G'Ä:R VA:R+ T"I:ÅNDE MIN'U:T
V'ÅLEON GN'Y:R BÂ:DEO+ D'Ä:G, Â+: N'ÄT
R"E:T'Ä "ALDR'IG "ANDR'AS J'U:R"Ä
FL"IK'AN B"A2S2T'ADE T"ÄNDÄ42NA "AFT'TA
TIL+ FR"UK'ÄST 'Ä:TÄ4R VI+: GR'Ö:T
B'IELENS "E:N'A J'U.L Ä+: TR"A:S'IT
SM"UTS'IGA SK'O:R SKA+ ST"ÄL'AS I:+ H'ALEN
H'ÄSTEN SKA:+ H"ÅL'AS I:+ STR"A:M'A T"Y:GL'AR
H'UNDEN BLE:V+ R'ÄD, Â+: J"ÖMD'E SEJ+
PL"A:"Ä:KL'ÄKOR "ING+G'Ä:R I:+ S"ÄM'ARENS FL"O:R'A
R'INGEN L'Å:G I:+ SM'UTSEN, Â+: BL"ÄNK'T'E
R"O:S'O2NA BL"OM'AR T"I:D'IT PA:+ V'Å:REN
G'ÄSEN S"YKL'AR TIL+ SK"O:LN ID'A:G
M"ORM'OR R'E:SÄ4R TIL+ ST"ÄK'ÅL'M ID'A:G
H'ÄMT'A T"I:DN'INGEN, SÄM+ L'IGÄ4R I:+ BR"E:VL'Ä:DAT
DÄM+ F"O42S2T'À A"HÄL'ÅNEN SMA:KAR B'ÄST
H"U:SON V'ÅNGNEN H"INDR'AR TRAF'I:KEN PA:+ S"ÄM'AREN
B'Ä:2NEN L'E:KÄ4R "U:TE PA:+ 'ANGEN
DE+ Ä+: H"ÄLSOSAMT AT+ DR"IK'A MJ'ÖLK