SPEECH READING DISCRIMINATION OF HOMOPHENES

G. A. LATTE

One particular area of research in speech reading relates to the analysis of confusions among visually perceived speech sounds. These visually confusable groups of sound are called 'Homophenes'. The few studies reported do not agree upon the findings with respect to the speech reading discrimination of homophenes, (Erber 1972$_b$ and Walden and Prosek 1974$_a$).

Erber (1972$_b$) studied the distinction between eight common consonants /p, b, m, t, d, n, k, i, in the bisyllabic context. Three groups of subjects namely, normal hearing, severely hearing impaired and profoundly deaf children were used. All three groups were able to distinguish between the classes of sounds but not within a class of sound.

It is generally accepted that it is not possible to distinguish between /p,b,m,/ and /t,d,n/ etc., homophenes purely by vision.

However, when Walden and Prosek (1974$_a$) attempted to train hearing impaired adults to distinguish visually within the homophenes categories Two weeks of training resulted in considerable improvement in this direction.

Thus, it appears that correct perception of confusable visemes within the same homophenous category may be possible because of certain subliminal cues or because of some obvious yet hitherto unsuspected cues.

The present study attempts to answer the question whether correct perception of homophenes is a matter of chance? Or whether there are some other cues yet known, which help in accurate speech reading discrimination of homophenes and therefore it is not a matter of chance ?

The hypothesis

The following null hypothesis was put to experimental test.

'Correct identification of homophenes within the same category is no more than a matter of chance'.

METHOD

(a) The subjects

Ten subjects (five males and five females) with normal vision and hearing, in the age of 17 years 9 months to 21 years 4 months (mean 19 years 2 months) participated in the study. None of them wore glassess and none reported of any visual problems. Their hearing was normal. Though their mother tongues differed, all the subjects had adequate knowledge of English

language in order to participate in the study. The subjects knew of speech reading but had no formal instructions in speech reading.

(b) The speakers

The young adult speakers, a male and a female spoke the stimulus material. The speakers had no speech defect and their competence in English language was adequate in order to serve as speakers in this study.

(c) The stimuli

The bilabials /p, b, m/ and the alveolars /t,d,n/ were the two homophene groups used in the study. CVC combinations were formed using the above consonants along with vowel /ae/ as a constant nucleus. The resulting monosyllables were such as PAD, NAB etc., when a bilabial sound occurred in initial position. The alveolar occurred in the final position and vice versa. Some of the monosyllables were meaningful words in English. A total of 18 such combinations were formed. Nine of these containing bilabials as the initial consonants were randomly assigned to each of the two speakers. Similarly the other nine containing the alveolars as initial consonants were randomly assigned to each of the two speakers who spoke these stimuli.

Recording of the stimuli

A video-tape (Type 1810 'U' maticsony) was used for recording the utterences along with the video-camera (Type AVC 3250 CE – Sony). A cardoid type microphone formed one of the parts of the entire system.

The speaker's ace was well illuminated by three intense lights (500 watts each). The camera was so placed that the speakers image appeared life size on the video monitor TV screen. Full face front view above the shoulders was recorded. The video screen was 18" wide and the picture was in black and white. The recording was done in sound treated regular recording studio.

The Testing

The subjects were first familiarized with the stimuli while video-tape ran with the sound switched on. Later, during actual testing the sound was switched off while the video-tape was replayed. The subjects speech read the stimuli as the image appeared on the video screen.

Each subject marked his/her own responses on a multiple choice type score-sheet. Each list was presented three times to permit the revision of responses.

Results and discussion

The results of the experiment are presented in table I to IV.

Table I and II show the per cent correct responses for male and female subjects respectively on various tasks.

Table 1

Sl. No. of subjects	Male speaker				Female speaker			
	Bilabials		Alveolars		Bilabials		Alveolars	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final
1	22.22	44.44	33.33	22.22	55.55	22.22	22.22	22.22
2	11.11	44.44	22.22	55.55	22.22	22.22	44.44	77.77
3	55.55	33.33	55.55	22.22	33.33	33.33	44.44	33.33
4	33.33	44.44	22.22	22.22	22.22	22.22	33.33	44.44
5	11.11	44.44	22.22	44.44	33.33	44.44	55.55	55.55
Mean	26.66	42.21	31.10	33.33	33.33	28.88	39.99	46.66
Minimum	11.11	33.33	22.22	22.22	22.22	22.22	22.22	22.22
Maximum	55.55	44.44	55.55	55.55	55.55	44.44	55.55	77.77

Table showing the percentage of correct responses obtained by the male subjects on various tasks

Table 2

Sl. No. of subjects	Male speaker				Female speaker			
	Bilabials		Alveolar		Bilabials		Alveolars	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final
6	22.22	33.33	11.11	22.22	22.22	44.44	33.33	44.44
7	22.22	33 33	33.33	55.55	44.44	11.11	33.33	0
8	44.44	22.22	44.44	44.44	22.22	33.33	44.44	77.77
9	33.33	44 44	44.44	33.33	44.44	11.11	55.55	66.66
10	100.00	44.44	22.22	66.66	22.22	11.11	22.22	44.44
Mean	44.44	35.55	39.99	44 44	31.10	22.22	37.37	46.66
Minimum	22.22	22.22	11.11	22.22	22.22	11.11	22.22	0
Maximum	100.00	44.44	44.44	66.66	44 44	44.44	55.55	77.77

Table showing the percentages of correct responses obtained by the female subjects on various tasks.

It can be observed that the mean per cent correct responses have not exceeded 46.66% for any consonant either in initial or in the final position. This suggests that the performance did not exceed chance. Six out of ten subjects exhibited random responses.

However, a few individuals have scored beyond chance. Scores such as 66.66%, 77.77% and 100%. Correct responses by some subjects (subjects No. 2, 8, 9, 10) in some tasks would suggest that their performances were not merely a chance.

Table III shows the criteria used to reject or accept the null hypothesis. The criteria are based on the binomial expansion theorem of probabilities. Number of correct responses such as zero ,one, six, seven, eight and nine etc., cannot be just a matter of chance according to this criteria. Table III also indicates the actual probability values of responses being due to chance and therefore tells whether to accept or to reject the hypothesis.

Table 3

No. of correct responses	Probability of being chance	
0	0.0260122	Reject
1	0.0130061	H_o
2	0.2341106	
3	0 2706539	Accept
4	0.2048468	H_o
5	0.1024234	
6	0 03414	Reject
7	0.00073159	H_o
8	0.0009144	
9	0.0000508	

Table showing the criteria used to reject or accept the null hypothesis, based on Binomial Expansion Theorem of probabilities.

Table IV shows the original scores of the ten subjects on the various tasks. Using the probability criteria from table III the responses that were probably not due to chance at 0.05 level of significance are underlined. It can be seen that one subject (subject 10) performed definitely on identification of bilabial visemes when speaker was a male. The correct responses are nine and their being due to chance is highly improbable. Subjects No. 2, 7, 8, 9, identified six times correctly the final alveolars. Such performances are not due to chance.

Table 4

Sl No.	Initial bilabial	Final bilabial	Initial alveolar	Final alveolar	Initial bilabial	Final bilabial	Initial alveolar	Final alveolar
1	2	4	3	2	5	2	2	2
2	1	4	2	5	2	2	4	7
3	5	3	5	2	3	3	4	3
4	3	4	2	2	2	2	3	4
5	1	4	2	4	3	4	3	5
6	2	3	1	2	2	4	3	4
7	2	3	3	5	4	1	3	0
8	4	2	4	4	2	3	4	7
9	3	4	4	3	4	1	5	6
10	1	4	2	6	2	1	2	4

Table showing the original scores by the ten subjects on various tasks.
The non-chance performances are underlined.

Rest of the subjects have showed purely chance performances on correct identification of visemes in the initial and final positions for both the bilabial and alveolar categories.

Several instances of beyond chance performance still leave the hypothesis open for further investigation and hence the null hypothesis that 'the correct perception of visemes among the same category is no more than a matter of chance' was rejected but with reservations.

Keeping in mind the fact that subjects were not formally trained in speech reading an alternative hypothesis that 'there seem to be some cues which help in speech reading discrimination of homophenes, but they are yet unknown' was accepted, however, with reservations. Some subjects correctly utilized these cues and some of them did not.

Implications of the study

The belief that it is not possible to discriminate within homophenous category purely by vision has placed a ceiling on the possibilities of proficiency that speech readers can be expected to attain.

But it appears that cues such as pressure in lip approximation for the bilabials and in the tongue contact for alveolars and / or the differences in time in articulation of different sounds and observable movements in laryngeal region during voicing etc., might be some of the cues that are available to speech readers. Such possibilities should be explored.

REFERENCES

Erber, N. P., 'Speech-Envelope cues as an acoustic Aid to Lip reading for profoundly Deaf Children', J. Acoust. Soc. Amer. 51, p. 1224-1227 (1972 b).

Walden, B. E. Prosek, R. A., and Worthington, D. W., 'Predicting Audio Visual consonant recognition performance of hearing impairment adults'. J. Speech Hear. Res. 17, p. 270-278 (1974 a).