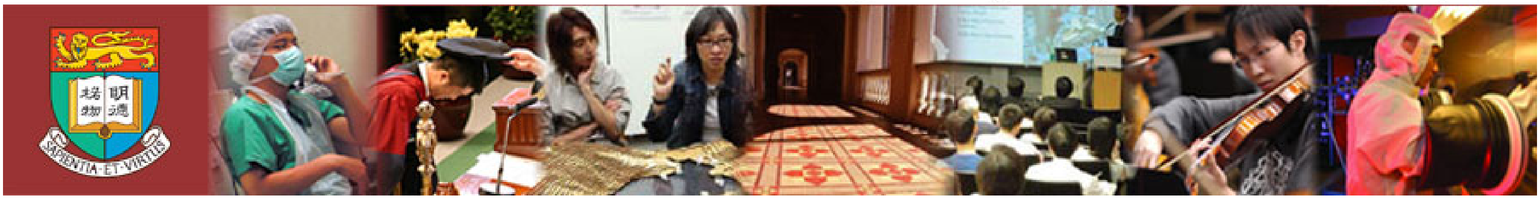


THE PERCEPTION OF CANTONESE ASPIRATED CONSONANTS AT DIFFERENT SIGNAL-TO-NOISE RATIOS

Wong Mun Yiu, Gladys

A dissertation submitted in partial fulfilment of the requirement of the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, April 30, 1992.

ABSTRACT

The present study examines the perception of five pairs of Cantonese aspirated and unaspirated consonants in Consonant-Vowel syllables under three different signal-to-noise ratios (S/N +15,0,-5). Fifteen female subjects (mean age 22.2) listened to 240 CV syllables overlapped by Cafeteria noise, 80 syllables under each S/N. The ten syllables were real words and printed on paper. The subjects responded to each item by circling one of the ten words on the paper. They sat four feet from a loudspeaker when listening to the syllables. The results show that aspirated consonants are more affected by the noise when they have the same intensity levels as the latter. Plosives are more difficult to recognize than clusters and affricates. Error analyses reveal that the error patterns change with S/N. Reasons for the confusion are discussed.

INTRODUCTION

The perception of English consonants has been studied extensively. Some studies on this topic are by Miller and Nicely (1955), and Fant (1973). Various studies investigated the discrimination of single words by hearing impaired subjects, such as the work done by Bess and Townsend (1977), Perkkarinen, Salmivalli & Suonpaa (1990), and Tyler (1982). Chermak and Dengerink (1984) studied the word identification in noise by normal hearing school children. All in all, much work has been done on related topics in English speaking population.

In contrast, Cantonese word discrimination in noise remains unexplored. This study, hence, aims at studying normal hearing adults' perception of aspirated consonants under different signal-to-noise ratios.

Miller and Nicely (1955) found that the place of articulation is more vulnerable to random masking noise than other features like nasality or voicing in English consonants. As aspiration is not a contrastive feature in English, it was not included in their study. The present study concentrates on this contrastive feature in Cantonese.

WHAT IS SIGNAL-TO-NOISE RATIO?

Signal-to-noise ratio (S/N) is the relative intensity of the signal (speech) to the background noise. For example, when the speech signal is 10dB stronger than the noise, the

S/N is +10. When they are equal in intensity levels, S/N is 0. (Martin, 1986) Hence, the lower the S/N, the more difficult it is to perceive the speech signal because the noise is louder than the speech. This study employed three S/Ns (+15,0,-5) by using noise at 45dBA through and varying the speech signals to 60, 45 and 40dBA.

WHAT ARE CANTONESE ASPIRATED CONSONANTS?

A contrastive feature gives a different meaning when a phoneme having this feature is replaced by another one without it or vice versa. (Hyman 1975) Aspiration is a contrastive feature in Cantonese. Take as an example the Cantonese words /p'au^{55}/ (THROW) and /pau^{35}/ (BREAD), they cannot replace one another, otherwise, the meaning changes. There are five pairs of aspirated consonants and their unaspirated counterparts in Cantonese. They are:

1) Aspirated bilabial plosive /p'/ e.g. /p'ai^{33}/ (distribute) and its unaspirated counterpart /p/ e.g. /pai^{33}/ (bow).

2) Aspirated alveolar plosive /t'/ e.g. /t'an^{55}/ (stall) and its unaspirated counterpart /t/ e.g. /tan^{55}/ (single).

3) Aspirated velar plosive /k'/ e.g. /k'im^{35}/ (clamp) and its unaspirated counterpart /k/ e.g. /kim^{35}/ (examine).

4) Aspirated alveolar affricate /ts'/ e.g. /ts'iu^{55}/ (super) and its unaspirated counterpart /ts/ e.g. /tsiu^{55}/ (banana).

5) Aspirated labiovelar cluster /k'w/ e.g. /k'wa^{55}/ (exaggerate) and its unaspirated counterpart /kw/ e.g. /kwa^{55}/ (vegetable).

These ten consonants are the targets in this study. Hashimoto (1972) classified the aspirated consonants as having the distinctive feature [tense]. According to Jakobson and Halle [1] (1956), the distinctive feature [tense] is characterized acoustically by a higher total amount of energy in conjunction with a greater spread of energy in spectrum and in time. Also, it is articulated with a greater deformation of the vocal tract away from its rest position. Since aspirated consonants have greater spread of energy in their frequency spectra, noise such as the Cafeteria which spreads widely in the spectrum may have a greater possibility masking them.

This study aims at answering the following questions: i) whether the noise will affect the discrimination of aspirated consonants; ii) whether the effect of noise will be greater on aspirated than unaspirated consonants; iii) whether different types of consonants will be differentially discriminable.

SUBJECTS

Fifteen female adults aged 18 to 25 (mean age 22.2)

---

[1] Jakobson and Halle (1956) is a frequent quotation in Hyman (1975). To save space, they are regarded as a separate reference in this paper.

participated in the experiment. Before proceeding to the test, their hearing were screened using pure tones of 0.5, 1.0 & 2.0kHz. The criteria of pass were set at 25 dBHL for each frequency. This screening procedure ensures that they have normal hearing. Otherwise, errors they made might be due to hearing loss.

MATERIALS AND INSTRUMENTATION

Selecting stimuli - any extraneous factors are controlled

The vowel /a/ following the five pairs of aspirated and unaspirated consonants formed CV syllables in the experiment. The phonetic context was limited to the single vowel, /a/, to control any extraneous factors due to vowel context change. Similarly, CV syllables were used instead of CVC ones to eliminate any possible effect of final consonants on initial ones. All the syllables had high level tones to avoid any differential effect due to tonal difference. They were real words in Cantonese. This rules out any potential discrepancies in discrimination due to nonsense syllables perception, since nonsense syllables are more difficult to recognize than real words. (Hodgson 1980) Each consonant appeared eight times in each S/N.

Recording stimuli

A man produced the ten target syllables with carrier phrases attached to each syllable. Recording the syllables with carrier phrases made them sound more natural. They were recorded on a cassette tape (Tape A). TDK SA-60ST cassette tapes were used through this experiment. Tape A was duplicated onto another tape (Tape B) at 3 different intensity levels (60, 45, 40dBA) measured by a sound level meter (Quest model 215). The choice of speech signal at 60dBA was appropriate since conversational speech is at about this level. While speech at 40dBA intensity was still audible to the subjects. Further reduction of the speech level would lead to great difficulty identifying the speech signal in noise, while increasing the speech level beyond 60dBA might cause discomfort. Given these unfavorable factors in further increasing or decreasing the speech signal, only three S/Ns were used in this study. Otherwise, the subjects might have low motivation participating in the test due to frustration or discomfort, hence the results would not be reliable.

Extracts of Tape B with the syllables (intensity 60, 45 and 40dBA) on then were stored in a NeXT computer for later recording.

Cafeteria noise was chosen in this study for mixing with the syllables because it is an everyday sound in modern daily life. Its frequencies spread from 0 to 8 kHz, with slightly greater amplitude between 4 to 8 kHz. The noise was recorded onto Tape B (from a compact disc <Widex Real Life Environment

Sound Examples>) at an intensity level 45dBA measured by the sound level meter. This noise was also stored in the NeXT and later cut down to a 2-second segment, because subjects might adapt to the noise if it was continuous.

The investigator then mixed the ten CV syllables with the 2-second noise in the computer. The noise at 45dBA overlapped syllables at 60, 45 and 40dBA. Hence three S/Ns (+15, 0, -5) were obtained. Then the syllables mixed with noise were recorded onto Tape C as stimuli to present to the subjects. There were 5-second silent intervals to give subjects time to respond. As each consonant had to appear eight times at one S/N, the recording procedure repeated eight times by using the same mixed syllables in the computer. They were randomly ordered during recording, so that the same syllable never follow itself.

After CV syllables from Tape B had been detached from the carrier phrases in the computer, they were recorded onto Tape D as well. They were the calibrating stimuli when administering the test.

PROCEDURES

Calibration procedures

The investigator preset three different intensity levels for delivering the syllables. The calibrating stimuli on Tape D were fed into the cassette recorder (ReVox B710 MKII) first, the volume (intensity) output was adjusted until 60dBA was measured by the sound level meter (from 4 feet) when the 60dBA calibrating stimuli were fed into the recorder. The knob control for volume output was marked. Repeating the above procedure gave intensity levels preset for speech signals at 45 and 40dBA as well. When delivering the stimuli, say at S/N +15, the investigator adjusted the volume output of the cassette recorder to the marked level for playing 60dBA calibrating stimuli. Hence, the maximum output was 60dBA only. The noise would be at 45dBA which was fixed during the recording procedure.

Testing procedures

After an audiological screening test, a subject sat 4 feet from a loudspeaker (Westra LAB-501) with an answer sheet. The ten target words were printed under each item with the same order on the answer sheet. The subjects were asked to read aloud the ten target words once before hearing any stimuli. This was to make sure that no errors were due to unfamiliarity with the target words. When a subject did not know a word, the investigator read it aloud and asked her to repeat it. Then when hearing the syllables, they chose one from the words. If they were uncertain, they were encouraged to make a guess. One syllable was presented at a time. Each subject listened to 80 syllables at each S/N, 240 syllables in all. Hence, it was a repeated measure design.

After finishing the first 80 syllables i.e. presenting syllables at one particular S/N, the investigator adjusted the volume of output (intensity level) to another preset level suitable for the succeeding S/N.

To reduce order effect on discrimination scores, three orders of S/Ns ([+15,0,-5], [0,-5,+15] and [-5,+15,0]) were used in the study.

Correct recognition of a target word scored one mark. Erroneous responses were recorded for further analysis. Scores of the aspirated and unaspirated consonants were summed separately to make up two separate scores.

RESULTS

Investigating the effect of S/N on aspirated and unaspirated consonants

Table 1 shows the mean and standard deviations of percentage recognition scores of aspirated and unaspirated consonants at different S/Ns. Differences were observed in scores of aspirated and unaspirated consonants when S/N changed. Moreover, their mean scores differed from one another under each S/N level . The effect of S/N on the discrimination scores of aspirated and unaspirated consonants was tested by the Friedman two-way ANOVA [2] respectively.

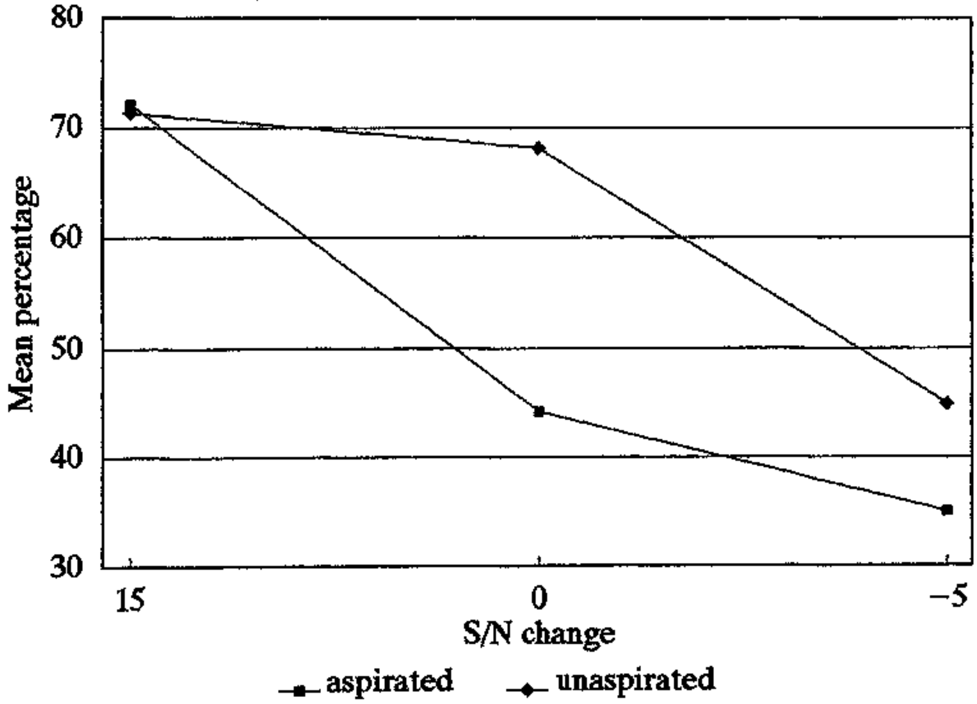
---

[2] In this paper, the ***nonparametric tests Friedman two-way analysis of variance by ranks, Kruskal-Wallis one-way analysis of variance by ranks and Wilcoxon matched-pairs signed rank test*** were applied since the data had marked heterogeneity of variance. However, these tests cannot always reveal where statistical differences, if any, lie. One, therefore, has to recourse to the descriptive statistics to indicate the difference.

Table 1 MEAN PERCENTAGE SCORES (n=40) AND STANDARD DEVIATIONS OF ASPIRATED AND UNASPIRATED CONSONANTS AT DIFFERENT S/N

	Aspirated		Unaspirated	
	mean	S.D.	mean	S.D.
S/N+15	72.17	7.89	71.33	12.48
S/N 0	44.18	10.23	68.18	11.87
S/N -5	35.00	10.40	44.83	11.27

Fig. 1 Mean percentage of aspirated and unaspirated consonants



There is a significant difference (observed value $.\chi_r^2..$ =26.5, ties excluded, [3]*df*=2, *p*<0.01) in score of aspirated consonants when S/N changes. The difference observed in the unaspirated consonants discrimination scores also reaches statistical significance (observed value $.\chi_r^2..$ = 19.9, ties excluded, *df*=2, *p*<0.01) due to the effect of S/N. Their scores, as revealed by Figure 1, are decreasing when S/N lowers.

Investigating the effect of aspiration on the discrimination scores

At each S/N level, the effect of aspiration on the percentage score was tested by the Wilcoxon matched-pairs signed rank test (see footnote 2). A significant difference is found (observed value of Wilcoxon statistic = 112, *p*<.05) between the percentage scores for aspirated and unaspirated consonants at the S/N 0 but not at +15 and -5 (observed values of the Wilcoxon statistic are 21 and 65 respectively, *p*>.05). This means that aspirated consonants have significantly lower percentage discriminated at S/N 0. Figure 1 illustrates this result. As S/N has significant effect on both the scores of aspirated and unaspirated consonants, and the effect of aspiration is significant on the discrimination

---

[3] The Friedman statistic ($.\chi_r^2..$) has a sampling distribution approximating the chi-square distribution with *df*=*k*-1 if the sample size is 9 or above when there are three conditions (*k*) or more under test. Also, chi-square distribution is applicable if there are more than 4 samples when *k*=4 or above. (Siegel 1956)

Table 2 The confusion matrix of errors at S/N +15

Response	Stimulus (n=120)									
	p'	p	t'	t	k'	k	ts'	ts	k'w	kw
p'	26	2	9		1					2
p	1	64	2	8						2
t'	57		64		17					
t		38	17	69		1				
k'	20	1	15	1	101				1	
k		14	12	40		78				
ts'					1	1	120			
ts	1		1	1		40		120		1
k'w	15	1							119	15
kw				1						100

Table 3 The confusion matrix of errors at S/N 0

Response	Stimulus (n=120)									
	p'	p	t'	t	k'	k	ts'	ts	k'w	kw
p'	25	3	4	6	18		2		2	1
p	5	70	11	32					5	
t'	57	1	4	14	67	1	3		1	
t	7	39	78	35	1	3				
k'	17			8	29		2			
k	5	6	23	23		77		5		
ts'					4		112	1		
ts	1			1	1	38		113		
k'w	3	1		1			1		98	7
kw						1		1	14	112

Table 4 The confusion matrix of errors at S/N −5

Response	Stimulus (n=120)									
	p'	p	t'	t	k'	k	ts'	ts	k'w	kw
p'	52	1	1	3	6				3	
p	5	23	6	25	6	4		2	25	
t'	22	3	13	1	29	3	21	2	3	
t	1	59	55	43	24	12		16	4	
k'	7	3	4		17	4		4	2	
k	3	31	40	39	29	68		58	5	1
ts'	3				4		96			
ts	1			3	1	22	3	37		
k'w	23			4	3			1	32	21
kw	3		1	2	1	7			46	98

scores at S/N 0, it indicates that both the aspirated and unaspirated consonants discrimination are affected by the S/N, in particular, aspiration leads to significant difference in accuracy at S/N 0.

The effect of type of consonants on the discrimination scores

Apart from the effect of aspiration, some other possibly pertinent variables are intrinsic to this study: the place and manner of articulation. Table 2 to 4 show the confusion matrices for the ten different consonants at each S/N (pooled across subjects). Errors mainly occur for identifying plosives at S/N +15 and then some errors are found for clusters and affricates at S/N 0 and -5. In order to analyze any significant difference, three percentage scores are obtained for type of consonant: plosives, affricates and clusters.

Table 5 shows the mean and standard deviations of percentage of each type of consonants discriminated under different S/Ns. The observed difference between each type of consonants under each S/N was tested by using the Friedman two-way ANOVA. The results show that type of consonant has significant effect on the percent scores at S/N +15 (χ_r^2=21.2, *df*=2, $p<.01$), S/N 0 (χ_r^2=23.7, *df*=2, $p<.01$) and S/N -5 (χ_r^2=14.7, *df*=2, $p<.01$), by referring to the chi-square distribution. The descriptive statistic shows that clusters and affricates have similar percentage scores and both these scores exceed plosives' across the S/Ns.

Table 5 MEAN PERCENTAGES AND STANDARD DEVIATIONS OF DIFFERENT TYPES OF CONSONANTS RECOGNIZED AT DIFFERENT S/N

	Affricates		Clusters		Plosives	
	mean	S.D.	mean	S.D.	mean	S.D.
S/N +15	100.00	0.0	90.83	12.7	56.61	13.2
S/N 0	93.75	8.9	87.08	13.2	32.92	11.0
S/N -5	55.42	21.9	54.17	15.8	30.14	8.8

The Friedman two-way ANOVA was used to test the effect of S/N on the discrimination scores of each type of consonants. The effect of S/N is significant on percentage scores of clusters (χ_r^2 =22.2, *df*=2, *p*<.01), affricates (χ_r^2 =21.9, *df*=2, *p*<.01) and plosives (χ_r^2 =19.3, *df*=2, *p*<.01). Their scores decrease with S/N.

Although the consonants recognized decrease with S/N, percentage of plosives discriminated is always less than that of affricates and clusters.

Testing the effect of S/N on the error distribution and the type of error on percentage of error

To reveal the error pattern, error analyses were done under all S/Ns. Errors were categorized into four types: (i) aspiration change, (ii) placement change, (iii) manner change, and (iv) others. Examples of each type of error follow:

Error type	Target	Response
Aspiration change	/p'a/	/pa/
Placement change	/ta/	/ka/
Manner change	/ts'a/	/t'a/
Others	/p'a/	/ta/

For the errors involving plosives only, classification is obvious. This is not so for clusters and affricates. For example, when /k'w/ is confused with /p'/, the error, here, is assumed to involve confusion in manner and place of articulation. However, it may not be the true picture. It is because they differ in both features. A listener may be able to perceive /k'w/ by either feature, since there is no labiovelar plosive or bilabial cluster. Single feature, therefore, may be sufficient to cause the error. However, one cannot tell which feature leads to the error. As a result, it is assumed that both features are confused.

Table 6 MEAN PERCENTAGE DISTRIBUTION OF DIFFERENT ERRORS ACROSS THE S/N

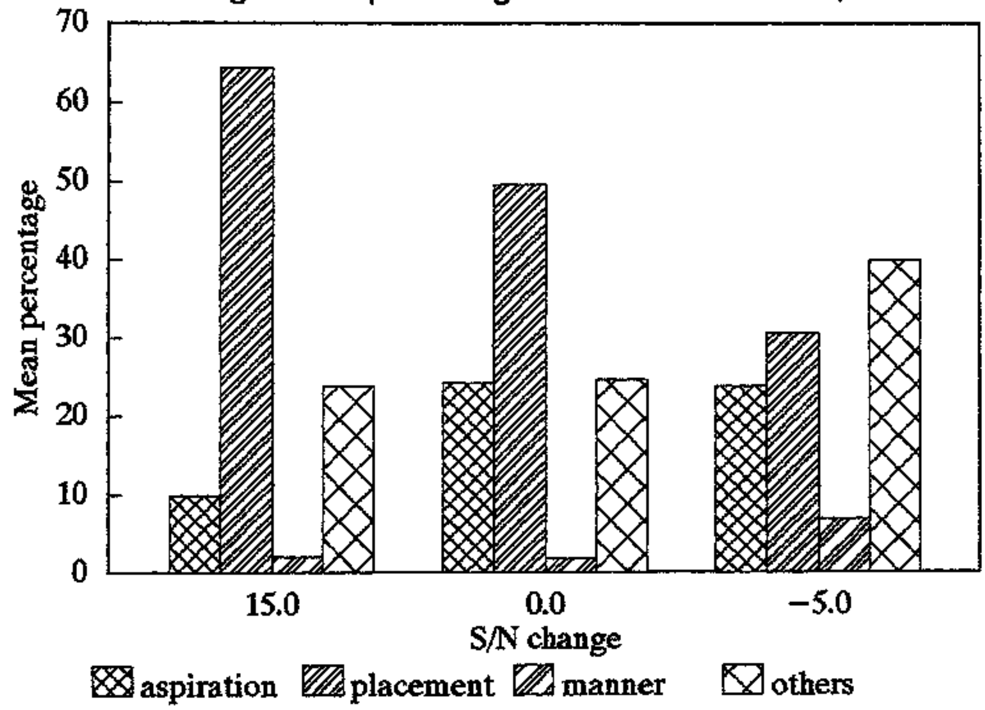
	Aspiration	Placement	Manner	Others
S/N +15	9.8	64.4	2.0	23.8
S/N 0	24.1	49.6	1.7	24.6
S/N -5	23.8	30.5	6.9	39.8

Table 6 shows that the errors are not evenly distributing at each S/N level. Friedman two-way analysis of variance by ranks was used to test whether the observed distribution bias reachs statistical significance at each S/N level. The results show that percentage distribution of errors does differ significantly by the error type at each S/N level. At S/N +15, the observed value χ_r^2=39.9, *df*=2, *p*<.01, while χ_r^2 equals 27.1, (*df*=2, *p*<.01) and 40.0 (*df*=2, *p*<.01) at S/N 0 and -5 respectively (by chi-square distribution). That is, the distribution of errors is not even at all S/N levels.

To analyze the effect of S/N on each error, Friedman statistic was done for four times, percentage of each error shows significant difference (*df*=2, *p*<.01) due to the effect of S/N. The observed value of χ_r^2=16.9 for "Aspiration",

χ_r^2 =22.9 for "Placement", χ_r^2 =9.5 for "Manner" and χ_r^2 =12.4 for "Others". The results indicate that all the errors have an increase in their percentages and this corresponds to a decrease in the "Placement error" when S/N falls. Figure 2 illustrates these findings.

Fig.2 Mean percentage of errors at different S/Ns



To investigate the order effect

There were three orders of presenting syllables at different S/Ns in this study, they are [+15,0,-5], [0,-5,+15] and [-5,+15,0]. The effect of different orders on discrimination scores of aspirated and unaspirated consonants was tested by the Kruskal-Wallis one-way ANOVA [4]. The observed *H* values at different S/Ns are summarised in Table 7.

Table 7 THE RESULTS OF KRUSKAL-WALLIS ONE-WAY ANOVA TESTING THE EFFECT OF DIFFERENT ORDER ON DISCRIMINATION SCORES OF ASPIRATED AND UNASPIRATED CONSONANTS AT EACH S/N LEVEL

	H value	*df*	*p* value
S/N +15			
aspirated	1.07	2	>.1
unaspirated	3.82	2	>.1
S/N 0			
aspirated	0.86	2	>.1
unaspirated	2.60	2	>.1
S/N -5			
aspirated	3.28	2	>.1
unaspirated	1.69	2	>.1

[4] When the number of conditions (*k*) is 3 and the sample is more than 5, the Kruskal-Wallis statistic (*H* value) can approximate chi-square distribution. Hence, chi-square distribution is not applicable here for the samples are not more than 5.

The effect of order is not significant ($p>.1$) in either the discrimination scores of aspirated or unaspirated consonants at any S/N level.

DISCUSSION

The findings support that aspirated consonants are more difficult to recognize than unaspirated consonants (at S/N 0). Clusters and affricates are easier to recognize than plosives, and the error pattern changes with S/N.

Why clusters and affricates are better recognized?

To explain the poorer recognition of plosives, one needs to examine their acoustic properties. According to Dorman & Hannley (1985), plosives are characterized by: i. a silent interval produced by vocal tract occlusion, then ii. there is a brief burst of energy at release of occlusion, after that, iii. frication is produced due to turbulence in the narrow but still opening space between articulators, and iv. an aspiration period follows before voicing onset, finally v. there is a formant transition when the articulators move from the point of occlusion to a different configuration for the following vowel. These characteristics are cues for the perception of a plosive. A listener will combine these cues, depending on the redundancy of them, to eliminate incorrect solutions and search for the correct one. To recognize a plosive correctly, the listener then has to perceive the differences among the plosives in the following aspects: i) duration, amplitude and spectrum of the burst energy, for example alveolar and bilabial plosives have greater spreads of energy than the velars. (Jakobson and Halle 1956). Aspirated plosives have even greater spread, longer duration of burst and greater energy compared with the unaspirated ones. ii) Different plosives have different formant transition durations. iii) They also have different formant onset frequencies and directions of formant movement. According to Fischer-Jorgensen (1954), the starting frequencies of formant transition are in descending order, alveolar, bilabial, velar. Velar and alveolar plosives have falling F2 transitions, and those of bilabials are rising. iv) They also differ in the voice onset time (VOT), i.e. the sum of release burst, frication and aspiration durations. The mean VOTs are, in ascending order, bilabial, alveolar, velar (Dorman et al 1977). The correct recognition of a plosive, therefore, requires complex feature comparisons, i.e. comparing energy amplitudes, burst spectra, formants (to discriminate the place of articulation), the silent intervals (to recognize the manner of articulation), and VOTs (for the manner and place of articulation as well). Given this complexity, it is reasonable that plosives have lower percentage recognition scores than other consonants. The reasons are: a) plosives' silent intervals might be masked by the noise. b) Intensities of the noise produced by plosives are less strong than others. Fry (1979) claimed that affricates had longer period of frication and greater intensity than plosives. Accordingly, we would expect that affricates are easier to recognize than plosives. c) Clusters have downward shift of upper frequencies in their spectra. (Jakobson and Halle 1956). This shift might be a significant cue in the noise. d) Plosives' formants change rapidly (Fry 1979). Dorman & Hannley (1985) quoted a finding of Yokkaichi & Fujisaki (1978) that plosives' formant transitions are brief, hence difficult to identify for hearing impaired subjects. This situation is probably similar for the normal hearing subjects in noise. In short, one understands the factors leading to lower percentage recognition scores in the plosives.

Unaspirated consonants are better recognized

The above explanation for lower percent scores of plosives does not seem to account well for the performance of aspirated consonants. According to the above argument, longer VOT and greater intensity of noise produced by the consonant, should favour the recognition of consonants. Aspirated consonants do have these properties, but they have lower recognition scores than the unaspirated ones at S/N 0. This difference should lie in other properties such as energy spectrum and formant transitions.

Miller and Nicely (1955) claim that as voiceless consonants have their formant transitions in the aspiration period, they are more difficult to hear. If this argument is correct, formant transitions of aspirated consonants will be more difficult to hear than unaspirated ones in the noise, because of their even greater aspiration period (Fischer-Jorgensen 1954). A significantly masked formant transition would lead to lower recognition scores in aspirated consonants. In addition, aspirated consonants have greater spread of energy, hence the noise which has energy spread over different frequencies may mask them more readily.

However, the observed difference here varies with S/N levels. No difference is found statistically between aspirated and unaspirated consonants when S/N level is +15 or -5. This variability is probably due to the acoustic properties of the noise and the aspiration feature. Noise is an aperiodic sound. For consonants the vocal tract generates aperiodic sounds, too. This noise becomes more irregular when the aspirated consonants are produced because they have a longer aspiration phase and larger intensity. (Fry 1979) When this noise segment of speech is presented against an aperiodic background noise of the same intensity, they will be difficult to discriminate. Indeed, when the background noise is much less strong than the speech signal, its masking effect is not stronger for either the aspirated or unaspirated consonants. The noise is also not differential in effect for aspirated and unaspirated consonants at S/N -5.

This may be because the noise produced by both is already masked well by the noise. As a result, less aspirated consonants are recognized than the unaspirated at S/N 0 only.

Error distribution changes with S/N - the dominant error changes

Differences between errors and within each type are significant at all S/N levels. Referring to figure 2, one sees that the proportion of errors changes with S/N. The dominant error is "Placement" at S/N +15 and 0 and replaced by "Others" (random errors) at S/N -5. "Placement" error is due to substitution of a target consonant by another which has the same manner and aspiration feature but differs in its place of articulation. The result indicates that the errors are mainly confusion of consonants of different places at S/N +15 and 0, i.e. only one feature is confused (the place of articulation). When S/N falls to -5, the nature of confusion between consonants no longer depends on a single feature but on multiple features. This implies that the cues for perception of correct consonants break down in a stepwise fashion. This explanation assumes that correct recognition of a consonant requires integrating three features, "Aspiration", "Place and manner of articulation". The noise first affects the "Place of articulation" with "Aspiration" and "Manner of articulation" grossly preserved. As a result, the consonants are recognized correctly for manner and aspiration, perhaps by integrating the latter two features. When S/N further falls to -5, the noise has effect on all the features and the integration process becomes further strained. The consonant confusion is no longer due to one feature disruption, i.e. the place of articulation only. Instead, manner and aspiration, or place and aspiration are confused together. The confusion becomes more nearly random and the errors accordingly more evenly distribute.

The present findings are not parallel to the findings of Miller and Nicely's work (1955). Because this study involves the aspiration feature and uses a different noise. On the contrary, Miller and Nicely employed noise at 200-6500 cycles per second which is an artificial noise not found in daily life. These factors might have differentiated the results.

Limitations of the study

The experimental procedures require the subjects to discriminate one consonant out of ten. This procedure has introduced more difficulty than needed if only the aspiration feature is studied. One procedure that might be better is to give binary choices. That is, only the aspirated consonant and its unaspirated counterpart are given. Hence, what the subject has to do is to discriminate between aspirated and unaspirated consonants, without the need to further discriminate the place and manner of articulation.

As mentioned in the result section, a problem arises in classifying errors which may lead to misinterpretation. This is the problem of classifying errors involving the clusters and affricates. Take as an example of the error process /kw -> p/, this error is classified as "Others" since one cannot tell whether the place or the manner is confused. Therefore, the error is assumed to be due to difficulty perceiving both features. This assumption may be wrong because the two features seem redundant in the perception process of the consonant /kw/. Thus, single feature may be sufficient to perceive the sound. If so, even when S/N is reduced to -5, the confusion may be due to difficulty perceiving a single feature instead of both. However, one assumes multiple feature confusion in interpreting the results. This assumption may need further consideration. This problem is due to introduction of more than one variable yet being systematic in this study.

The study resulted in less than 100% accuracy in the consonants discrimination even at a S/N level of +15. This might be due to errors or an early masking effect of noise at S/N +15. The former possibility could be eliminated if there were a control experiment in that the syllables were delivered without noise. Discrimination approximating 100% could rule out this possibility.

FURTHER STUDIES

The experimental design introduced some variables which were not systematic. For example, place error is only applicable to plosives, similarly manner error is applicable to /ts', ts and t', t/ only. This distribution might have biased the errors. In further studies of the aspiration and place of articulation, it might be useful if one focused on one or two pairs of aspirated and unaspirated consonants, such as /t', t, p', p/.

Since the interpretation of the present findings incorporated some acoustic data on English or Danish consonants, acoustic analyses of Cantonese consonants would be useful for interpreting the results on the perception of Cantonese speech sounds.

CONCLUSION

The study confirms that aspirated consonants are affected by the S/N, and this effect holds for unaspirated consonants as well. When the noise was at the same intensity as the speech signal, aspirated consonants discrimination was worse than unaspirated consonants. The corresponding difference was not significant when the speech signal was greater or smaller in intensity than the noise.

Plosives are more difficult to recognize than affricates or clusters. Although the three groups were recognized less well when the speech signal was attenuated, plosives still had the lowest recognition scores across the S/Ns.

Error patterns did change across the S/Ns. The

"Placement" error predominated at S/N +15 and 0, while the "Others" errors replaced it at S/N -5. This indicates that the consonants are mixed up due to single feature confusion initially, then multiple feature confusion becomes the source of error. The correct perception of a consonant becomes more difficult when all its features are affected by the noise.

ACKNOWLEDGEMENT

I would like express my gratitude to Professor Zubrick, Dr. Harrison, Dr. Ciocca and Mr. Au, who gave me their suggestions at various stages of my study, and the technicians who always offered help in handling the equipment, and all the people who participated in the experiment or gave comments on this paper.

REFERENCES

Bess, F.H. & Townsend, T.H. (1977) Word discrimination for listeners with flat sensorineural hearing losses. *Journal of Speech and Hearing Disorders*, **42**, 232-237.

Chermak, G.D. & Dengerink, J. (1981) Word identification in quiet and in noise. *Scandinavian Audiology*, **10**, 55-60.

Dorman, M.F. & Hannley, M.T. (1985) Identification of speech and speechlike signals by hearing impaired listeners. In Daniloff, R.G. (Ed) *Speech Science*. California: College-Hill Press.

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

Fant, G. (1973) *Speech sounds and features*. Massachuetts: MIT.

Fischer-Jorgensen, E. (1954) Acoustic analysis of stop consonants. *Miscellanea Phonetica*, 2, 42-59.

Fry, D.B. (1979) *The physics of speech*. Cambridge: Cambridge University Press.

Hashimoto, O.K.Y. (1972) *Studies in Yue Dialects I: Phonology of Cantonese*. London: Cambridge University Press.

Hodgson, W.R. (1980) *Basic audiologic evaluation*. Baltimore: Williams and Wilkins.

Hyman, L.M. (1975) *Phonology: Theory and analysis*. New York: Holt, Rinehart and Winston.

Jakobson, R. & Halle, M. (1956) Fundamentals of language. 's-Gravenhage. (Quotations) In Hyman, L.M. (1975) *Phonology: Theory and analysis*. New York: Holt, Rinehart and Winston.

Martin, F.N. (1986) *Introduction to Audiology*. Englewood Cliffs: Prentice Hall.

Miller, G.A. & Nicely, P.E. (1955) An analysis of perceptual confusions among some English consonants. Reprinted in Saporta, S. (Ed.) (1961) *Psycholinguistics: A book of readings*. New York: Holt, Rinehart and Winston, 153-175.

Perkkarinen, E., Salmivalli, A. and Suonpaa, J. (1990) Effect of noise on word discrimination by subjects with impaired hearing, compared with those with normal hearing. *Scandinavian Audiology*, **19**, 31-36.

Pickett, J.M. (1980) *The sounds of speech communication: A primer of acoustics phonetics and speech perception*. Texas: PRO-ED, Inc.

Siegel, S. (1956) *Nonparametric statistics for the behavioral sciences*. Tokyo: McGraw-Hill Inc.

Tyler, R., Wood, E. and Fernandes, M. (1982) Frequency resolution and hearing loss. *British Journal of Audiology*, **16**, 45-63.