# THE GENDER AND INDIVIDUAL VARIATIONS IN PROCESSING LINGUISTIC-PHONETIC CUES

U Thein-Tun

Department of Communication Disorders
La Trobe University

ABSTRACT - The perception of integrated phonetic cues by males and females was investigated at five levels of information processing. The integrated phonetic cues were the intensity and duration of voice-onset-time in relation to the intensity of the following vowel for syllable initial /d/-/t/ distinction. The five levels of information processing were auditory, phonetic, syllable, word, and sentence levels. The results demonstrate that listeners who cannot effectively process the cues at the auditory and phonetic levels can process them very effectively at the sentence level and vice versa. Most female listeners belong to the former group.

## INTRODUCTION

Virtually every phonetic contrast is cued by several distinct acoustic properties of the speech signal. within limits set by the relative perceptual weights and by the ranges of effectiveness of cues, the phonetic percept is maintained even after a change in the setting of one cue if it is offset by an opposed change in the setting of another cue. This phenomenon is generally known as a phonetic cue-trading relation (Fitch, Halwes, Erickson and Liberman, 1980). The distinct acoustic properties constituting one phonemic contrast are generally known as integrated cues. The majority of psychophysical research in speech perception conducted during the last four decades seems to be based on one assumption, that is, the manner in which individuals process a set of integrated cues in an isolated syllable or word context is the same as the way they process the same set of cues in the connected speech context. This assumption in fact is yet to be ascertained. In a connected speech context such as a sentence, prosodic features such as rate, rhythm, Fo pattern as well as syntactic and semantic information would certainly influence the integrated cue perception for phoneme identification. However, if these variations at the sentence level are controlled, would the integrated cue perception at the sentence level be the same as that at the lower levels such as auditory, phonetic and syllable levels for all individuals and males and females? It is the general objective of the present investigation to provide some answers to this question. Thein-Tun (1987) reported that there is an overall statistically significant difference in cue-trading relationship for the /d/ -/t/ contrast between males and females. The specific objective of this paper is to quantitatively describe the gender and individual differences across the five levels of information processing.

## METHOD*1

### Selection of segments and creation of linguistic levels

The acoustic properties investigated as integrated cues for the initial /d/ - /t/ contrast were the aspiration duration (VOT) and the aspiration amplitude in relation to the amplitude of the following vowel. These acoustic components were placed at five different linguistic levels of information processing, namely, auditory, phonetic, syllable, word and sentence levels. The VOT duration at each level varied from zero to 45 ms in ten 5-ms steps. This ten step VOT continuum may be described for each linguistic level as follows:

Auditory level: sinewave analogue of /da/ to /ta/.
Phonetic level: same as the auditory level
Syllable level: synthetic speech /da/ to /ta/
Word level: synthetic speech /dai/ "dye" to /tai/ "tie"
Sentence level: same as the word level but with stimuli contained in carrier sentences

The alveolar stop consonant voicing contrast was selected for investigation as alveolar stops have the most confined range of initial F2 and F3 frequencies (Fant, 1973, p. 124). The vowel /a/ was chosen for the syllable level continuum as its F2, the most important for consonant recognition, is in the lowest frequency range compared to its counterparts in other vowels. Listeners' perception is known to be more stable in the low frequency region (e. g. Pickett et al., 1972). The diphthong /ai/ was chosen for the word level continuum to provide suitable lexical items. The stimuli of the present investigation had to belong to five different linguistic levels. At each level they had to carry not only the same phonetic contrast but also the

same phonetic context so that any interlevel variation that may be found in the results would not be influenced by the difference of the phonetic context in which the stimuli occurred. The first steady state of the diphthong /ai/ represents the closest possible phonetic environment to /a/ while /ai/ as a vocalic segment is capable of forming a word either with /d/ or /t/. At the sentence level, in order to create the sentence level processing, the individual steps of the /dai/-/tai/ continuum were placed at the end of carrier sentences which had low predictability (PL) of their last word. The creation of the syllable level stimuli will be described first since they served as a link between the two lower and two upper levels.

The syllable level stimuli

Using a 12 parameter serial analogue speech synthesiser designed by Clark (1976), and a synthesis time frame of 5 ms, the first step of the syllable level VOT continuum , that is the /da/ end of the /da/-/ta/ continuum (VOT=0 ms), was created. The frequency values of the three formant patterns of this synthetic /da/ were determined from averaged measurements of the five /da/ spectrograms of five male native speakers of general to broad Australian English. The VOT continuum for the remaining nine 5-ms steps was created by replacing the periodic voiced (V) excitation with noise and simultaneously increasing the band-width of the F1 transition to its maximum and hence virtually eliminating the existence of the F1 transition. The first /da/-/ta/ continuum created in this way produced a good /da/ at one end and a good /ta/ at the other. The amplitude levels of the noise and vowel portions in this continuum (though different in actual measurements) were given the reference values of 0 dB. Therefore the first /da/-/ta/ continuum can be described as bearing the nominal amplitude pattern of 0 dB A (aspiration noise) and 0 dB V (vowel portion). Eight more /da/-/ta/ continua were created by increasing and decreasing both the A = 0 dB and V = 0 dB amplitudes by 6 dB as described below.

| A amplitude | | V amplitude |
|---|---|---|
| +6 dB | orthogonally | +6 dB |
| 0 dB | combined | 0 dB |
| -6 dB | with | -6 dB |

Each of the A values was combined with each of the V values to produce nine amplitude environments.

Four randomised presentations of ten VOT steps in nine amplitude environments were made. A 360 stimulus tape was produced with an interstimulus interval of 3.5 seconds. These four sets of randomised stimuli served as four blocks of test stimuli for the syllable level.

Phonetic and auditory level

As mentioned earlier, the phonetic and auditory level stimuli were the sinewave analogues of the syllable level stimuli. They were arranged on the test tape in four randomised blocks of 90 stimuli, as for the syllable level stimuli. For the phonetic level test, the subjects were told that the stimuli were the whistled imitation of the /da/ and /ta/ speech sounds. For the auditory level test, the subjects were told that the stimuli were non-speech computer sounds; they were also instructed to treat the sinewave analogue of /da/ as "sound one", and the sinewave analogue of /ta/ as "sound two".

Word level stimuli

The word level stimuli were in principle the same as the four blocks of 90 stimuli at the syllable level. The difference was that the word level stimuli were from the nine dai/-/tai/ continua instead of the nine /da/-/ta/ continua of the syllable level. With the exception of the initial transition duration, the durations and frequency values of the formant trajectories in the good /dai/ were determined from the corresponding averaged values of the five male native speakers of general to broad Australian English. As with the /da/ end stimulus at the syllable level, the duration of the initial formant transition was 45 ms in the /dai/ end stimulus. Such an arrangement was necessary in order to maintain the uniformity of VOT steps at different linguistic levels. It was the duration of the initial formant transitions which was progressively replaced by noise in order to create the VOT steps in these experiments.

Sentence level stimuli

The sentence level stimuli were the same four blocks of the word level /dai/-/tai/ ("dye" and "tie" as words) stimuli. In order to create the sentence level processing for the subjects, the "dye"-"tie" stimuli had to appear in the sentence context. Therefore carrier sentences had to be formulated. In formulating the carrier sentences the followoing criteria were followed in order to control the semantic and syntactic

variations and vocabulary of the sentences:

(i) The stimulus words (dye/tie/) appeared at the end of the sentence following the article "the";
(ii) The message of the carrier sentence predicted the final word with low probability ;
(iii) The carrier sentences were constructed using three syntactic components of subject, predicate and object. The stimulus word filled the object slot with the article. The predicate slot consisted of t wo monosyllabic auxiliary verbs ("should have", "might have", "may not" etc.) and one monosyllabic head verb. Hence the predicate slot consisted of three syllables. The subject slot was filled by a phonetically simple disyllabic proper noun. Every sentence therefore consisted of seven syllable s;
(iv) No alveolar stop (/d/ or /t/) in the syllable initial position was to be included in the carrier sentences.

On the sentence level test tape, the 90 carrier sentences were synthesised according to a synthesis by rule system of Australian English (Clark 1981). Spectrograms processed from 14 carrier sentences (7 carrying the /dai/ token and 7 carrying the /tai/ token) spoken with the same rate and intonation by the same speaker (a trained phonetician) were used as norms in synthesising the carrier sentences. The average amplitude of the carrier sentences was kept at the same value as the nominal 0 dB of the V amplitude of the stimulus word.

Subjects, test setting and procedure

Thirteen male and fifteen female normal hearing listeners participated in the listening tests. They were all native speakers of Australian English and naive listeners of synthesised speech, aged between 20 and 40 years. The listening tests were conducted in an acoustically treated speech perception laboratory. The tape output level was adjusted in such a way that the amplitude of loudest sound on the tape was approximately 80 dB SPL. An anchoring procedure was used to familiarise the subjects with the end points of the stimulus continuum before the actual listening test for each level. At the four lower levels, the task of the subjects was to identify every stimulus, whereas they were required to write down every sentence as they heard it at the sentence level. Each subject was tested at each of the five levels with at least two weeks time lapse between tests at different levels. At each level each subject was tested with all four randomised blocks. Therefore every subject had a possible score of up to four /d/ (or "sound one" at the auditory level) or four /t/ (or "sound two" at the auditory level) responses for each of the 90 stimuli (3 A levels X 3 V levels X 10 VOT steps). See Thein-Tun (1987) for details.

DATA ORGANISATION

In order to examine the roles of the A and V amplitudes in each of the nine continuua, firstly the nine original amplitude patterns were regrouped under two headings, namely , first grouping and second grouping.

| First grouping<br>For the V factor analysis (A constant and V varying)<br>original continuum No. | A/V amplitude pattern | | combined pattern |
|---|---|---|---|
| 1 | A = +6 dB / V = +6 dB | | |
| 2 | A = +6 dB / V = 0 dB | A constant at +6 dB | 1 |
| 3 | A = +6 dB / V = -6 dB | V varying | |
| 4 | A = 0 dB / V = +6 dB | | |
| 5 | A = 0 dB / V = 0 dB | A constant at 0 dB | 2 |
| 6 | A = 0 dB / V = -6 dB | V varying | |
| 7 | A = -6 dB / V = +6 dB | | |
| 8 | A = -6 dB / V = 0 dB | A constant at -6 dB | 3 |
| 9 | A = -6 dB / V = -6 dB | V varying | |

While the purpose of the first grouping was the V factor analysis, the purpose of the second grouping was the A factor analysis (V constant and A varying), the reverse of the first grouping. In the combined pattern 1 of the first grouping, A is constant at +6 dB while V varies at the three different levels. Therefore examining the responses of combined pattern 1 of the first grouping is in effect examining the varying V amplitude factor while A is constant at +6 dB. Similarly, examining the responses of the

same phonetic context so that any interlevel variation that may be found in the results would not be influenced by the difference of the phonetic context in which the stimuli occurred. The first steady state of the diphthong /ai/ represents the closest possible phonetic environment to /a/ while /ai/ as a vocalic segment is capable of forming a word either with /d/ or /t/. At the sentence level, in order to create the sentence level processing, the individual steps of the /dai/-/tai/ continuum were placed at the end of carrier sentences which had low predictability (PL) of their last word. The creation of the syllable level stimuli will be described first since they served as a link between the two lower and two upper levels.

The syllable level stimuli

Using a 12 parameter serial analogue speech synthesiser designed by Clark (1976), and a synthesis time frame of 5 ms, the first step of the syllable level VOT continuum , that is the /da/ end of the /da/-/ta/ continuum (VOT=0 ms), was created. The frequency values of the three formant patterns of this synthetic /da/ were determined from averaged measurements of the five /da/ spectrograms of five male native speakers of general to broad Australian English. The VOT continuum for the remaining nine 5-ms steps was created by replacing the periodic voiced (V) excitation with noise and simultaneously increasing the band-width of the F1 transition to its maximum and hence virtually eliminating the existence of the F1 transition. The first /da/-/ta/ continuum created in this way produced a good /da/ at one end and a good /ta/ at the other. The amplitude levels of the noise and vowel portions in this continuum (though different in actual measurements) were given the reference values of 0 dB. Therefore the first /da/-/ta/ continuum can be described as bearing the nominal amplitude pattern of 0 dB A (aspiration noise) and 0 dB V (vowel portion). Eight more /da/-/ta/ continua were created by increasing and decreasing both the A = 0 dB and V = 0 dB amplitudes by 6 dB as described below.

| A amplitude | | V amplitude |
|---|---|---|
| +6 dB | orthogonally | +6 dB |
| 0 dB | combined | 0 dB |
| -6 dB | with | -6 dB |

Each of the A values was combined with each of the V values to produce nine amplitude environments.

Four randomised presentations of ten VOT steps in nine amplitude environments were made. A 360 stimulus tape was produced with an interstimulus interval of 3.5 seconds. These four sets of randomised stimuli served as four blocks of test stimuli for the syllable level.

Phonetic and auditory level

As mentioned earlier, the phonetic and auditory level stimuli were the sinewave analogues of the syllable level stimuli. They were arranged on the test tape in four randomised blocks of 90 stimuli, as for the syllable level stimuli. For the phonetic level test, the subjects were told that the stimuli were the whistled imitation of the /da/ and /ta/ speech sounds. For the auditory level test, the subjects were told that the stimuli were non-speech computer sounds; they were also instructed to treat the sinewave analogue of /da/ as "sound one", and the sinewave analogue of /ta/ as "sound two".

Word level stimuli

The word level stimuli were in principle the same as the four blocks of 90 stimuli at the syllable level. The difference was that the word level stimuli were from the nine dai/-/tai/ continua instead of the nine /da/-/ta/ continua of the syllable level. With the exception of the initial transition duration, the durations and frequency values of the formant trajectories in the good /dai/ were determined from the corresponding averaged values of the five male native speakers of general to broad Australian English. As with the /da/ end stimulus at the syllable level, the duration of the initial formant transition was 45 ms in the /dai/ end stimulus. Such an arrangement was necessary in order to maintain the uniformity of VOT steps at different linguistic levels. It was the duration of the initial formant transitions which was progressively replaced by noise in order to create the VOT steps in these experiments.

Sentence level stimuli

The sentence level stimuli were the same four blocks of the word level /dai/-/tai/ ("dye" and "tie" as words) stimuli. In order to create the sentence level processing for the subjects, the "dye"-"tie" stimuli had to appear in the sentence context. Therefore carrier sentences had to be formulated. In formulating the carrier sentences the followoing criteria were followed in order to control the semantic and syntactic

variations and vocabulary of the sentences:

(i) The stimulus words (dye/tie/) appeared at the end of the sentence following the article "the";
(ii) The message of the carrier sentence predicted the final word with low probability ;
(iii) The carrier sentences were constructed using three syntactic components of subject, predicate and object. The stimulus word filled the object slot with the article. The predicate slot consisted of t wo monosyllabic auxiliary verbs ("should have", "might have", "may not" etc.) and one monosyllabic head verb. Hence the predicate slot consisted of three syllables. The subject slot was filled by a phonetically simple disyllabic proper noun. Every sentence therefore consisted of seven syllable s;
(iv) No alveolar stop (/d/ or /t/) in the syllable initial position was to be included in the carrier sentences.

On the sentence level test tape, the 90 carrier sentences were synthesised according to a synthesis by rule system of Australian English (Clark 1981). Spectrograms processed from 14 carrier sentences (7 carrying the /dai/ token and 7 carrying the /tai/ token) spoken with the same rate and intonation by the same speaker (a trained phonetician) were used as norms in synthesising the carrier sentences. The average amplitude of the carrier sentences was kept at the same value as the nominal 0 dB of the V amplitude of the stimulus word.

Subjects, test setting and procedure

Thirteen male and fifteen female normal hearing listeners participated in the listening tests. They were all native speakers of Australian English and naive listeners of synthesised speech, aged between 20 and 40 years. The listening tests were conducted in an acoustically treated speech perception laboratory. The tape output level was adjusted in such a way that the amplitude of loudest sound on the tape was approximately 80 dB SPL. An anchoring procedure was used to familiarise the subjects with the end points of the stimulus continuum before the actual listening test for each level. At the four lower levels, the task of the subjects was to identify every stimulus, whereas they were required to write down every sentence as they heard it at the sentence level. Each subject was tested at each of the five levels with at least two weeks time lapse between tests at different levels. At each level each subject was tested with all four randomised blocks. Therefore every subject had a possible score of up to four /d/ (or "sound one" at the auditory level) or four /t/ (or "sound two" at the auditory level) responses for each of the 90 stimuli (3 A levels X 3 V levels X 10 VOT steps). See Thein-Tun (1987) for details.

DATA ORGANISATION

In order to examine the roles of the A and V amplitudes in each of the nine continuua, firstly the nine original amplitude patterns were regrouped under two headings, namely , first grouping and second grouping.

| First grouping<br>For the V factor analysis (A constant and V varying)<br>original continuum No. | A/V amplitude pattern | | combined pattern |
|---|---|---|---|
| 1 | A = +6 dB / V = +6 dB | | |
| 2 | A = +6 dB / V = 0 dB | A constant at +6 dB | 1 |
| 3 | A = +6 dB / V = -6 dB | V varying | |
| 4 | A = 0 dB / V = +6 dB | | |
| 5 | A = 0 dB / V = 0 dB | A constant at 0 dB | 2 |
| 6 | A = 0 dB / V = -6 dB | V varying | |
| 7 | A = -6 dB / V = +6 dB | | |
| 8 | A = -6 dB / V = 0 dB | A constant at -6 dB | 3 |
| 9 | A = -6 dB / V = -6 dB | V varying | |

While the purpose of the first grouping was the V factor analysis, the purpose of the second grouping was the A factor analysis (V constant and A varying), the reverse of the first grouping. In the combined pattern 1 of the first grouping, A is constant at +6 dB while V varies at the three different levels. Therefore examining the responses of combined pattern 1 of the first grouping is in effect examining the varying V amplitude factor while A is constant at +6 dB. Similarly, examining the responses of the

left hand panel
varying V amplitude levels

right hand panel
varying A amplitude levels

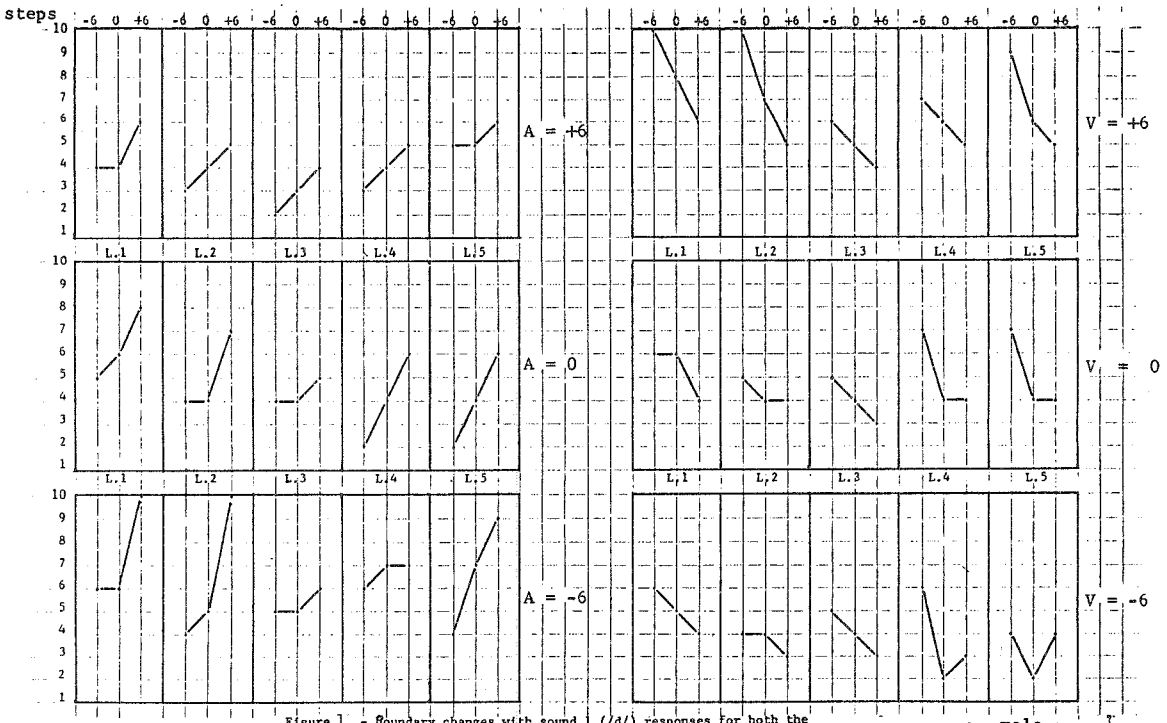


Figure 1 - Boundary changes with sound 1 (/d/) responses for both the A and V factors along the five linguistic levels

SUBJECT - NH 10 male