THE EFFECT OF LEXICAL STATUS ON THE PERCEPTION OF TONE

Robert A. Fox & James Unkefer
Ohio State Univeristy

ABSTRACT:

Chinese and American subjects were requested to identify speech tokens from four different Tone 1—Tone 2 *continua*. *These continua represented a word/word* continuum ([FEI1]—[FEI2]), a word/nonword continuum ([HEI1]—[HEI2]), a *nonword/word* continuum ([SHEI1]—[SHEI2]) and a nonword/nonword continuum ([DEI1]—[DEI2]) in Mandarin. Results showed that the tonal category boundaries of Chinese subjects were shifted in the direction of the nonword endpoints for both the word/nonword and nonword/word continua relative to the word/word continuum. The American subjects showed no such shift. These data were interpreted as suggesting that the lexical status of a speech token may significantly affect phonetic processing.

1. INTRODUCTION.

Recently there has been much interest in understanding the influence of "higher" levels of linguistic knowledge upon phonetic processing. For the most part, this research has been directed at determining the extent to which semantic and/or syntactic information may affect the categorization of individual phonetic segments. Results from various studies have demonstrated that nonphonetic contextual information may significantly affect the perception of somewhat ambiguous phonetic segments. For example, Isenberg, Walker, Ryder and Schweickert (1981) demonstrated that the phoneme boundary of a [t^{h}ə-ðə] continuum could be shifted in the direction of either endpoint depending on the nature of the surrounding grammatical context. In particular, the perception of the initial consonant varied significantly depending on whether the following word was a noun (which produced more 'the' responses) or a verb (which produced more 'to' responses). Similar effects have been reported by Garnes and Bond (1976) for a [b-d-g] place of articulation continuum and by Miller, Green and Schermer (1982) for a [b-p] VOT (Voicing Onset Time) continuum.

Other studies have shown that the lexical status of the stimulus token alone, unbiased by surrounding context, could produce significant variations in the perception of the token's initial segment. For example, a study by Rubin, Turvey and Van Gelder (1976) found that phoneme detection was faster when the target segment appeared in a word than in a nonword. Ganong (1980) demonstrated that the perception of ambiguous segments from related VOT Continua could differ significantly depending on whether or not the tokens of which they were part produced real English words. Fox (1982a, 1982b, 1984) showed that this lexical effect extends to a place-of-articulation continuum ([b-d]) and, in addition, found some evidence suggesting that word frequency variations could also produce phoneme boundary shifts.

Given these data on segmental perception, a logical question would be, is the perception of tone, or suprasegmental variations in general, affected in a similar fashion by lexical status distinctions? It is clear that language background (i.e., linguistic knowledge at a variety of levels) can influence the results of psychophysical experiments utilizing tone stimuli. For example, Wang (1976) conducted an experiment with both Chinese and American subjects using a stimulus set composed of an rising-tone—level-tone continuum for the syllable [i]. There were two tasks: identification and discrimination. Results from this experiment indicated American subjects produced data only consistent with psychophysical phenomena (i.e., the easiest stimulus to distinguish from the rest was the real level tone), while the Chinese subjects were affected to varying degrees by the linguistic (phonemic) boundary between the rising and level tones. A quite different type of support for the contention that tone perception can be influenced by linguistic background can be seen in the crosslanguage study of tone perception by Gandour and Harshman (1978)—a multidimensional scaling study of tone perception using speakers of Thai, Yoruba and English. The effect of linguistic background (and/or experience) was reflected in the differential emphasis placed on selected perceptual dimensions obtained during INDSCAL analysis. Both Gandour (1981) and Zhang et al. (1982) came to similar conclusions in studies using Chinese subjects only.

The present study was designed to address the issue of interaction betwen tone perception in Mandarin Chinese and linguistic information of a particular type, namely lexical status. In particular, the experiment to be described was designed to determine if the perception of ambiguous items in correlated tone continua can be shifted as a function of whether or not the stimulus item produces a real word.

2. METHODOLOGY.

2.1. SUBJECTS.

Eleven Chinese subjects and eleven American subjects took part in the listening test. Since the number of readily-available subjects who were native Mandarin speakers having a similar linguistic backgrounds was extremely limited (and certainly less than the minimum number needed to conduct this experiment) the choice was made to use subjects with a similar linguistic background but who had Mandarin (with native competence) as a second language. These subjects were from northern Taiwan and spoke the Min dialect as their native language. They learned Mandarin as a second language in elementary school. The American subjects all had English as their native language and none had studied a tone language of any type. The American subjects were included in the experiment primarily as a control group. All subjects were Ohio State University students and were paid volunteers. None had any known hearing impairments.

2.2. TEST STIMULI.

Mandarin Chinese has been recognized as having four distinctive tones for the citation syllable. Of these four tones, the high-level (Tone 1) and mid-rising (Tone 2) were selected for the experiment. Since these two tones have approximately the same F_0 level at the end-point and the frequency of F_0 rises in a linear fashion for the mid-rising tone, it was possible to produce Tone 1—Tone 2 continua by varying the starting F_0 frequency in discrete steps and linearly interpolating the F_0 over the duration of the syllable. To test for the lexical effect on tone perception, four different pairs of tokens were selected to give all possible combinations of word-nonword orderings. Each pair consisted of a Tone 1 token and a Tone 2 token. For all four token pairs the vowel was the same, EI , although the initial consonants differed for each continuum. The possible percepts are shown in Table 1 and each is marked for its word or nonword status. The tokens were determined to be a word or nonword by reference to various Chinese dictionaries (e.g., Lin, 1972), Liu's (1973) word frequency dictionary, and judgments by four native speakers of Mandarin.

Table 1: Lexical status of stimulus tokens.

	Stimulus Continuum			
	FEI	HEI	SHEI	DEI
High Tone	Word 'not' 非	Word 'black' 黑	Nonword	Nonword
Mid-Rising	Word 'fat' 肥	Nonword	Word 'who' 誰	Nonword

The test stimuli were created by producing a Tone 1-Tone 2 continuum for each of these token pairs. These continua were synthesized using the Klatt software cascade/parallel synthesis program (Klatt, 1980, 1978) implemented on the Ohio State University Linguistics Department's PDP 11/23. The F_0 values used in creating these tone continua were based on measurements by Howie (1970). The tokens varied in a step-like manner along a Tone 1-Tone 2 continuum. Each of the continua represented a set of 9 CV tokens which varied in terms of starting F_0 as shown in Figure 1. For each stimulus token F_0 rose in a linear fashion from this initial frequency value to 160 Hz after 200 msec. The F_0 for all tokens then linearly fell to 150 Hz at the end of the token. This contour created more natural sounding stimuli which were more representative of the contours described in Howie (1970). The voiced portion of all test stimuli (across all 4 continua) was 280 in duration. The change in initial F_0 between each adjacent stimulus was 4 Hz.

Representative values of the other acoustic parameters used in the synthesis of these tokens appear in Table 2. The vowel portion of all tokens (excluding 25 msec of consonant-vowel transition) was identical (except, of course for F_0 variations) for all tokens and was 250 msec in duration. The only non-suprasegmental differences among the 4 continua were those acoustic parameter values necessary to synthesize vowel initial [H], [F], [SH] and [D]—the actual parameter values used being modifications of those suggested by Klatt (1978). The three continuant consonants of ([H], [F], and [SH]) were generated by an aperiodic noise source filtered by

Table 2. Acoustic parameter values used in the synthesis of stimulus tokens.

	F1	B1	F2	B2	F3	B3	F4	B4	F5	B5
			Vowel EI (identical for all stimuli)							
Initial	480	70	1720	100	2520	200	3300	250	2850	200
Final	330	55	2020	100	2600	200	3300	250	3850	200
			Initial Continuant Consonants							
[F]	340	200	1100	120	2080	150	3300	250	3850	200
[H]	480	70	1720	100	2520	200	3300	250	3850	200
[SH]**					2750	300	3300	250	3850	200
			Formant Transitions for Initial Stop Consonant (D)							
Start	300	60	1700	100	2576	170	3300	250	3850	200
End	480	70	2020	100	2520	200	3300	250	385	200

*Fn represents formant frequency values; Bn formant bandwidths. All measurements in Hz.

**The fricative [sh] was produced through the parallel rather than the cascade circuit. Only the amplitudes of F3, F4 and F5 were set above 0 dB.

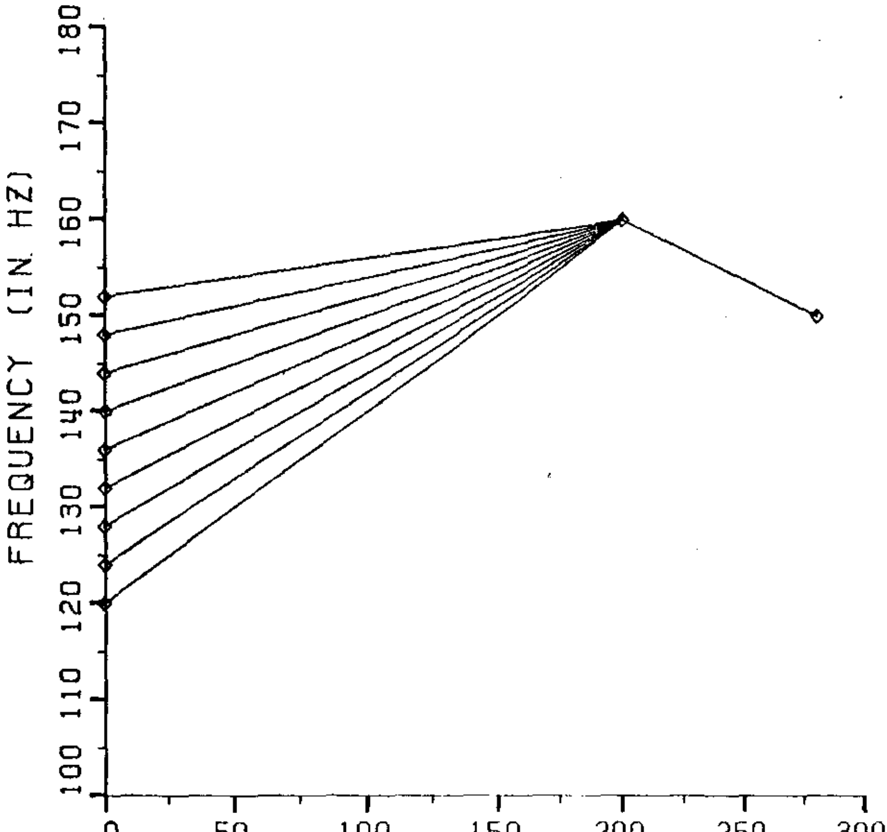
the appropriate resonance (formant) pattern. The formant frequency values used in constructing filter (transfer) functions are outlined in Table 2. In addition, there were formant transitions from these values to those of the following vowel [EI]. The fricative consonant durations were 100, 100 and 120 msec for [H], [F] and [SH], respectively. The initial stop consonant ([D]) was produced with a 10 msec burst followed by a 25 msec formant transition into [EI]. The starting and ending values of the transition are shown in Table 2.

Before the identification test was run the stimuli were evaluated by four Mandarin speakers to ensure that no obvious anomolies were present. These speakers were asked to judge the tokens for naturalness and to make word/nonword judgments. All tokens were judged to be relatively natural and the speakers agreed with the word/nonword classification of the tokens. Eight psuedo-random orderings of the 36 tokens were generated giving a total of 280 test tokens.

2.3. PROCEDURES.

The instructions for the Chinese subjects were written in Chinese and the subjects listened to ten tokens in order to set the volume level and be certain that the subjects understood the task. For the American subjects the instructions were in English and each participated in a short training session before the actual identification test. This training session was designed to

FIGURE 1. F0 CONTOURS OF 9 STIMULI



acquaint them with speech stimuli which varied in terms of F_0. During this session subjects were presented with 10 examples of the Tone 1 and Tone 2 endpoint stimuli for each of the four different test continua.

The identification task was a forced-choice design in which the subject had to determine whether the stimulus item was a Tone 1 or Tone 2 token, and respond by circling a "1" or "2" on the answer sheet. In addition, the subjects were asked to make a reliability judgment as to how certain they were of their choice by circling an "A", "B" or "C" to indicate complete confidence, likelihood, or a guess, respectively. This identification procedure is a variant of that used by Sawusch and Nusbaum (1979).

3.0. RESULTS.

The data from the first random ordering of stimulus items were treated as practice runs for all subjects. These data were eliminated from further analysis for all subjects, both Chinese and English. The identification plus confidence-rating responses were converted into a 6-point scale with a rating of "1" indicating a very confident Tone 1 identification and a "6" indicating a very confident Tone 2 identification. Note that although we will be discussing phoneme boundaries and evaluative statistics based only on these converted data, statistics based on the identification responses alone (i.e. % Tone 1 vs % Tone 2 responses) yielded practically identical results.

The identification responses for the Chinese group was summed over all subjects and are presented in Figure 2. The three identification functions shown are the two word/nonword continua and (HEI and SHEI) and the word/word continuum (FEI). The identification functions for the nonword/nonword continuum (DEI) were somewhat anomolous for both the Chinese and American subjects and will be discussed separately. Category boundaries for all four continua were calculated for each subject and are presented in Table 3. Note that there was a shift in the tone boundary in both HEI and SHEI toward the nonword boundary relative to the word/word continuum, i.e., there are relatviely more word responses than nonword responses. A one-way analysis of variance done on the data showed a significant effect due to tone continuum ($F(2,30)=3.98$, $p<.03$). Duncan's multiple range test indicated that there was a significant difference (at the .05 level) between the category boundaries for the HEI and SHEI continua. These results suggest that lexical information is having a significant effect upon suprasegmental categorization. However, since there are segmental differences among the tokens (i.e., initial consonant differences) that is perfectly correlated with the lexical differences among the continua these results cannot be accepted with the lexical effect interpretation until the responses from the American control group have been examined.

FIGURE 2.

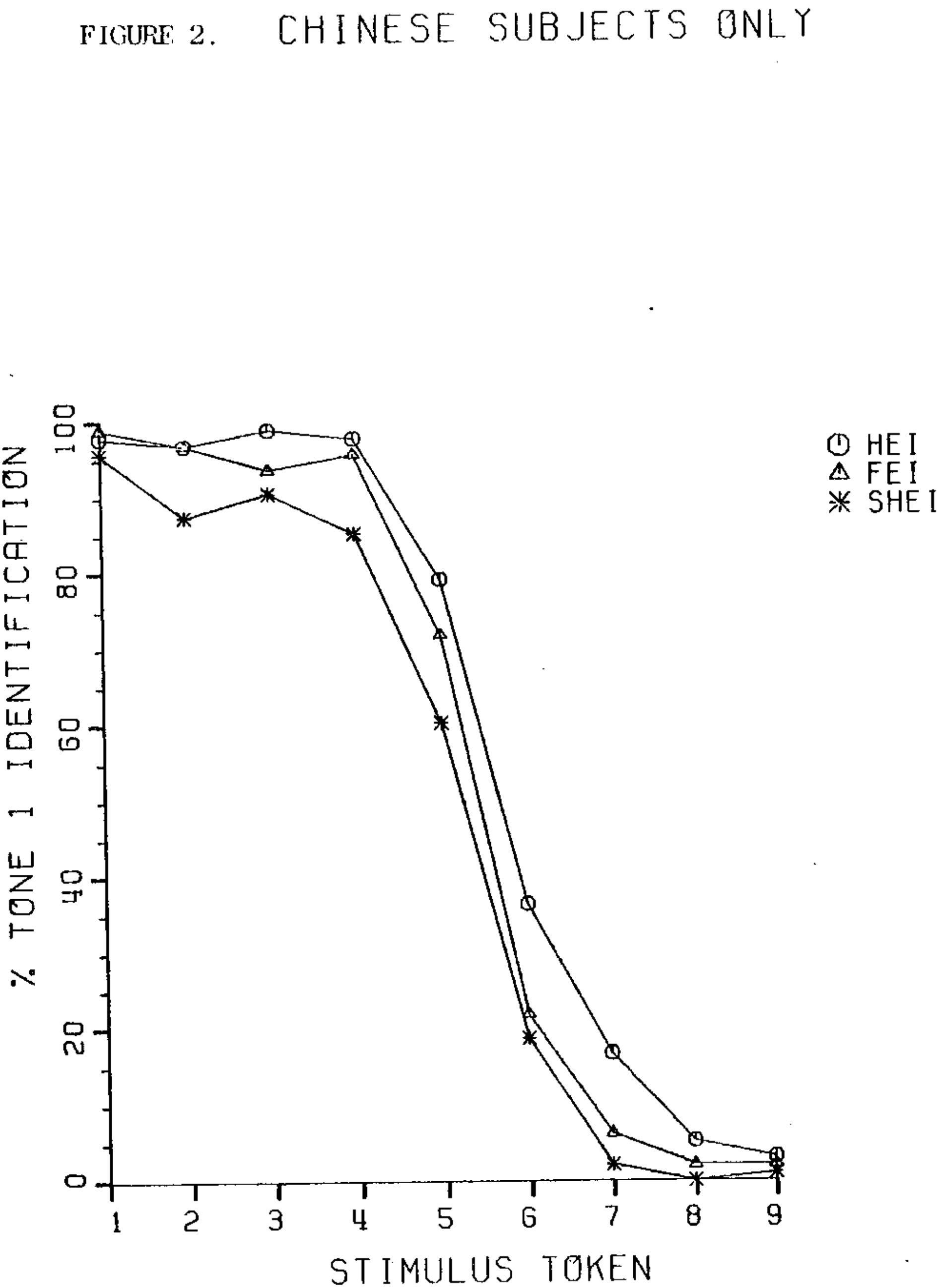
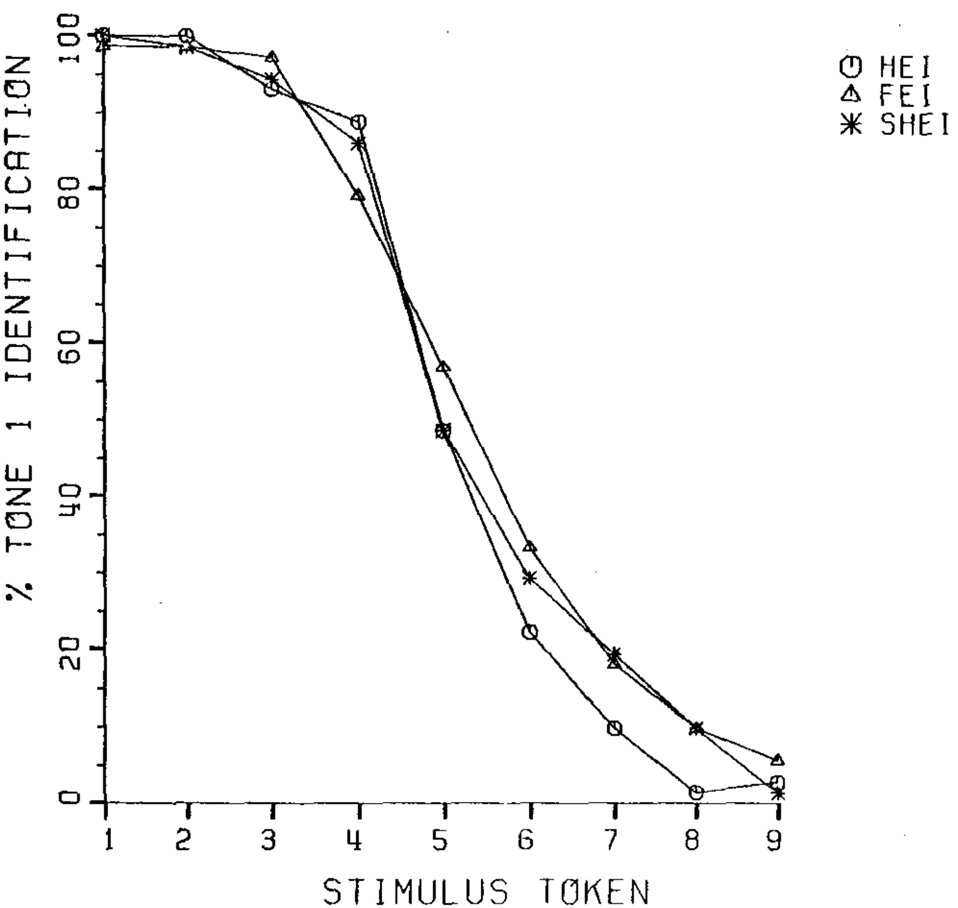

CHINESE SUBJECTS ONLY
% TONE 1 IDENTIFICATION
0
20
40
60
80
100
1
2
3
4
5
6
7
8
9
STIMULUS TOKEN
HEI
FEI
SHEI

FIGURE 3. ENGLISH SUBJECTS ONLY



The identification responses for the American control group are shown in Figure 3, and the appropriate tone boundaries appear in Table 4. It is obvious merely from an inspection of Figure 3 that the English group responses are not following those of the Chinese group. A one-way analysis of variance done on the English data show no significant main effect due to stimulus continuum ($F(2,30)=0.04$, n.s.). Duncan's multiple range test found no differences among the continua means at the .05 level or beyond.

The hypothesis that lexical status affects the suprasegmental categorization would predict that the nonword/nonword continuum would have a similar category boundary to the word/word continuum since subjects would not be biased toward either continuum endpoint in either case. This pattern of results was, in fact, found in Fox (1982a, 1982b, 1984). However, this particular prediction was not borne out in the present study. Indentification responses to the DEI stimuli were anomolous when compared to the other three continua. In particular, the phoneme boundaries for both the Chinese and American subjects were much closer to the Tone 2 endpoint than any of the other continua. The estimated category boundaries were significatly different from each of the other three continua for both the Chinese and American subjects (at the .05 level, Duncan's multiple range test). Since the phoneme boundary anomoly was true of both the Chinese and American

Table 3: Phoneme Boundaries--Chinese Subjects only.

Subjects	Stimulus Continuum			
	FEI	HEI	SHEI	DEI
1	5.35	7.88	4.38	6.50
2	4.74	5.27	5.25	6.05
3	4.79	5.72	5.73	9.00
4	5.48	5.92	5.73	6.50
5	5.40	4.87	4.63	5.65
6	5.21	5.25	5.14	5.55
7	6.22	6.43	6.03	6.23
8	5.09	5.05	4.85	5.42
9	5.50	5.27	5.56	6.23
10	5.50	5.97	5.16	8.25
11	5.91	6.56	4.25	6.70
Mean	**5.38**	**5.83**	**5.05**	**6.55**

Table 4: Phoneme Boundaries—American Subjects only.

Subjects	Stimulus Continuum			
	FEI	HEI	SHEI	DEI
1	6.17	6.68	7.27	7.45
2	3.59	4.50	5.03	4.63
3	5.70	5.58	5.42	7.83
4	4.81	4.45	3.83	5.37
5	5.38	4.97	4.59	6.54
6	5.36	4.89	5.25	6.27
7	5.28	4.47	5.19	6.43
8	6.75	6.95	6.20	8.22
9	3.53	4.27	4.47	4.68
10	5.79	5.34	5.17	6.54
11	6.75	6.32	6.58	8.68
Mean	**5.37**	**5.26**	**5.34**	**6.60**

subjects this result was not a simple function of the nonword nature of both endpoints (which the American subjects were not cognizant of). One might consider the possibility that the DEI anomoly, for the Chinese subjects, might arise as a function of language interference with Taiwanese (their first language). However, this does not seem to be the case. In a post-test interview each subject was asked to give a character for each of the 8 continua endpoints where possible, this was done to ensure that each subject had the predicted word/nonword classifications. No subject gave any response to the nonword tokens other than to state that no such word existed and no Taiwanese words were suggested. Post-experiment follow-up interviews with these and other Taiwanese speakeres concerning possible Taiwanese word responses for the Mandarin nonword tokens were also negative.

There are a number of nonlexical factors which could have contributed to the pattern of DEI responses. In the post-test debriefings subjects identified the stop consonant as the least "natural" sounding of the stimulus segments and perhaps this fact contributed to its uncharacteristic identification function. In addition, since the [D] was generated by using only a 10 msec stop burst and a 25 msec transition, the [D]-tokens were on the order of 65-85 msec shorter than the other stimulus tokens. There is also the possiblitiy that the perception of the DEI continuum was influenced by segmental factors. As the relevant phonetic literature suggest, the fundamental frequency of a vowel may be lower when the vowel follows a voiced segment than when it follows a voiceless one, all other things being equal (Hombert, 1978; Umeda, 1981). In particular, Umeda (1981) found that average peak F_0 for vowels was lower when preceded by d than when preceded by [f], [s] or [h], for two American English speakers. Seen from a perceptual viewpoint, if listeners somehow expect this type of F_0 pattern (i.e., relatively lower F_0 following [d] than [f], [s] or [h]) then when determining relative (linguistic) F_0 they might compensate for this phonetic tendency by utilizing a different frequency criterion by which to determine whether the F_0 is mid or high at vowel onset. Although such an account is speculative, it would explain the pattern of DEI responses in both the English and Chinese subjects. That is, since they normally expect a lower F_0 after [d], subjects perceive the onset F_0 of each step of the DEI continuum as relatively higher than the equivalent steps in the FEI, SHEI and HEI continua and the Tone 1-Tone 2 phoneme boundary would be nearer the Tone 2 endpoint for the DEI continuum than for the other 3 continua.

4. DISCUSSION AND CONCLUSIONS.

These results demonstrate that the perception of tone, a suprasegmental, can be affected by lexical status variation. In particular, the category boundary for tone continua which have one endpoint representing a word and the other a nonword are shifted toward the nonword endpoint, relative to a neutral word/word continuum. These data, together with those of Ganong (1980) and Fox (1984) suggest that the processing of both segmental and suprasegmental information can be modified by nonphonetic, linguistic information. However, determining just where in a model of perception to locate this effect is another question. Ganong (1980) argues that the lexical status effect operates simultaneously with or prior to the process of phonetic categorization itself. In particular, he argues that it is an instance of higher level linguistic knowledge directly interacting with the processing of auditory information. However, Fox (1982a, 1982b, 1984) argues that the lexical effect must manifest itself as a post-categorization "correction" process and that the requirement that phonetic categorization interact with top-down information was incompatible with his reaction-time data. The question as to the level of processing at which this lexical effect occurs with regard to the perception of tone variations is currently being investigated by the present authors using Fox's (1984) reaction-time design.

BIBLIOGRAPHY

FOX, R.A. 1982a. The effect of lexical status on identification of ambiguous speech tokens. Paper read at the 44th Meeting of the Midwestern Psychological Association.

------. 1982b. Influence of lexical status on phonetic categorization. Paper read at the 103rd Meeting of the Acoustical Society of America.

------. 1984. Effect of lexical status on phonetic categorization. Journal of Experimental Psychology: Human Perception and Performance. 10. 526-540.

GANDOUR, J. 1981. Perceptual dimensions of tone: evidence from Cantonese. JCL. 9. 20-36.

GANDOUR, J. and R. Harshman. 1978. Crosslanguage differences in tone perception: a multidimensional scaling investigation. Language and Speech. 21. 1-33.

GANONG, W. 1980. Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance. 6. 110-125.

GARNES, S. and Z.A. Bond. 1976. The relationship between semantic expectation and acoustic information. In Dressler, Pfeiffer and Herok eds., Phonologica 1976. Akten der dritten internationalen Phonologie-Tagung.

KLATT, D. 1978. Analysis and synthesis of CV syllables in English. Unpublished manuscript.

------. 1980. Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America. 67. 971-995.

HOMBERT, J-M. 1978. Consonant types, vowel quality, and tone. In Fromkin ed., Tone: A Linguistic Survey. Academic Press. New York.

HOWIE, J.M. 1976. Acoustical Studies of Mandarin Vowels and Tones. Cambridge University Press.

ISENBERG, D., E. Walker, J. Ryder and J. Schweickert. 1981. A top-down effect on the identification of function words. Paper read at 101st Meeting of the Acoustical Society of America.

LIN, Y. 1972. Chinese-English Dictionary of Modern Usage. The Chinese University of Hong Kong.

LIU, E. 1973. Frequency Dictionary of Chinese Words. Mouton. The Hague.

MILLER, J.L., K. Green and T. Schermer. 1982. On the distinction between prosodic and semantic factors in word identification. Paper presented at the 103rd Meeting of the Acoustical Society of America.

RUBIN, P., M.T. Turvey and P. Van Gelder. 1976. Initial phonemes are detected faster in spoken words than in spoken nonwords. Perception and Psychophysics. 19. 621-637.

SAWUSCH, J.R. and H.C. Nusbaum. 1979. Contextual effects in vowel perception I: Anchor-induced contrast effects. Perception and Psychophysics. 25. 292-302.

UMEDA, N. 1981. Influence of segmental factors on fundamental frequency in fluent speech. Journal of the Acoustical Society of America. 70. 350-355.

WANG, W. S-Y. 1976. Language change. Annals of New York Academy of Sciences. 61-72.

ZHANG, J., L. Shinan and Q. Shigian. 1982. A cluster analysis of the perceptual features of Chinese speech sounds. JCL. 10. 190-206.