Journal of Phonetics (1990) **18**, 481–495

Integration of segmental and tonal information in speech perception: a cross-linguistic study

Bruno H. Repp

Haskins Laboratories, New Haven, CT, U.S.A.

Hwei-Bing Lin

Graduate Center, City University of New York, NY, U.S.A.

Received 28th December 1989; and in revised form 25th April 1990

For speakers of a tone language, a close functional association exists between segmental structure and F_0 contour (i.e., tone) in speech because both dimensions are needed to identify words. Using the speeded classification paradigm, which does not require lexical access, we examined the hypothesis that segmental and tonal dimensions are perceptually more strongly integrated for speakers of a tone language (Mandarin Chinese) than for speakers of a nontone language (English). In four classification tasks, requiring attention to one dimension (either segmental or tonal) of CV syllables while ignoring the other, *both* subject groups showed strong perceptual integrality (i.e., interference from orthogonal variation in the unattended dimension). The Chinese subjects showed significantly more integrality than the English subjects in only one of the four tasks. After correcting for the fact that subjects took longer to respond to tonal distinctions than to segmental distinctions, we interpreted the results as suggesting that Chinese and English listeners both show an underlying processing asymmetry between consonants and tones in CV syllables, whereas only Chinese listeners show such an asymmetry between vowels and tones (vowels being more integral with tones than *vice versa*). This may be a reflection of specific language experience.

1. Introduction

Speech has a dual structure. At the articulatory level, there are two largely independent mechanisms: the laryngeal system which generates a relatively undifferentiated stream of sound whenever it is in action, and the supralaryngeal articulators which shape the sound into phonetic structures. In engineering terms, these systems may be characterized as (principal) source and filter (Fant, 1960). In the resulting acoustic signal, the effects of filter variation are largely restricted to frequencies above 200 Hz, whereas the effects of source (fundamental frequency, F_0) variation are most evident at lower frequencies, though they pervade the whole spectrum. In auditory perception, source variations appear as low-frequency, tonal

0095–4470/90/040481 + 15 $03.00/0

events, whereas filter variations appear as timbral changes. Typically, the rate of change is slower for F_0 changes than for filter changes. Finally, the functional load of these two constituents in the language is very different: while the filter variation is mainly responsible for the segmental structure of speech, the source variation is a principal carrier of information about intention, attitude and emotion.

How do listeners deal with this bipartite structure of speech? Do their brains separate the two types of information and process them in different centers? Or is speech processed as a single, integral structure? Two facts are beyond doubt. First, leaving aside certain subtle interactions, listeners can easily achieve a *cognitive* separation between source and filter information: they can make independent judgments about *how* some utterance was said ("questioning", "angrily", "by a child") *vs. what* it was ("really", "yes"). Second, listeners require both types of information for effective communication, so there is no *need* to separate them during processing. Different processing mechanisms may nevertheless have evolved in response to the different acoustic and informational characteristics of the two speech components. In that connection, it is relevant to consider that the source function is phylogenetically much older than the elaborate filter function of speech. Many aspects of source variation, especially as they relate to emotional states, have analogs in animal communication, whereas the filter function is almost uniquely human.

Indirect evidence bearing on the perceptual separation or integration of filter and source, or segmental and tonal, variation comes from neuropsychological investigations of hemispheric function. These studies have furnished considerable evidence that the left hemisphere of right-handed individuals is superior to the right hemisphere in the extraction and linguistic interpretation of segmental information. Some of this evidence comes from individuals with damage to the left hemisphere who, in addition to aphasia, commonly show impaired phonemic perception (e.g., Basso, Casati & Vignolo, 1977; Blumstein, Baker & Goodglass, 1977; Square-Storer, Darley & Sommers, 1988), whereas patients with damage to the right hemisphere or to nonspeech areas of the left hemisphere do not show such impairments (Basso *et al.*, 1977). Additional relevant data derive from dichotic listening studies of normal subjects that commonly show a right-ear advantage for consonantal contrasts, which has traditionally been interpreted as evidence for a left-hemisphere superiority (Kimura, 1967; Studdert-Kennedy & Shankweiler, 1970). Finally, there is also empirical support for a left-hemisphere superiority in the processing of auditory temporal information (e.g., Divenyi & Robinson, 1989; Halperin, Nachshon & Carmon, 1973), which may underly the left hemisphere's superior performance in processing the sequential phonetic structure of speech.

The right hemisphere, on the other hand, seems to outperform the left in the processing of constant or slowly varying pitch and spectral information (Divenyi & Robinson, 1989). Correspondingly, several studies have shown right-hemisphere advantages in normal subjects for the processing of speech intonation, using dichotic presentation of conflicting linguistic or affective intonation contours (Blumstein & Cooper, 1974; Ley & Bryden, 1982; Shipley-Brown, Dingwall, Berlin, Yeni-Komshian & Gordon-Salant, 1988). There is also considerable clinical evidence for prosodic disturbances in speech production following right-hemisphere damage (see Edmondson, Chan, Seibert & Ross, 1987).

It has been suggested that segmental and tonal information of speech are

processed independently in separate hemispheres (Goodglass & Calderon, 1977). At least one dichotic study of aphasic and normal individuals (Hartje, Willmes & Weniger, 1985), however, found evidence for good recognition of intonation contours by the left hemisphere. There are some data, moreover, suggesting that linguistic intonation evokes less of a left-ear advantage than does affective intonation (Shipley-Brown *et al.*, 1988), and that introduction of an appropriate prosody enhances the right-ear advantage for consonantal contrasts (Zurif & Mendelsohn, 1972). Shipley-Brown *et al.* (1988) hypothesized that, the stronger the linguistic function of some prosodic dimension, the more will it tend to be processed in the left hemisphere.

This issue becomes particularly interesting in tone languages such as Chinese or Thai, where, in addition to conveying affect, intonation serves the highly linguistic function of distinguishing lexical items. Indeed, not only do aphasic speakers of tone languages show an impairment in tone production (Naeser & Chan, 1980; Packard, 1986; Gandour, Petty & Dardarananda, 1988) but they also have difficulties with tone perception (Naeser & Chan, 1980; Gandour & Dardarananda, 1983). Dichotic studies of Van Lancker & Fromkin (1973, 1978) have shown that tonal distinctions evoke a right-ear advantage in normal speakers of a tone language (Thai), whereas the same stimuli produce no ear advantage in English-speaking listeners. In some sense, therefore, there is a closer tie between segmental structure and F_0 variation in tone languages than in nontone languages such as English. Chao (1980, p. 41) noted that "speakers of Chinese, whether literate or illiterate, feel unconsciously that a tone is such an integral part of the word that a syllable of the same consonantal and vocalic makeup but spoken in a different tone sounds like a totally different word".

Our intention in this study was to provide some behavioral evidence in support of Chao's observation, using a task that does not require lexical access. The question was: are segmental and tonal features, because of their constant association in lexical access, *perceptually* more integral for Chinese speakers than for English speakers? Results showing differences between these subject groups in hemispheric dominance for tonal (but not for segmental) distinctions do not necessarily imply differences in the separability of tonal and segmental features in real-time perceptual processing. Dimensions that are processed in the same hemisphere need not be perceptually integral, and dimensions processed in different hemispheres need not be separable, because of constant communication between the two halves of the intact brain. The question of the relative integrality or separability of stimulus dimensions during perceptual processing is more directly addressed by means of the *speeded classification paradigm* introduced by Garner (1970, 1974; Garner & Felfoldy, 1970), which we chose to employ in our experiment.

Garner's task requires subjects to pay attention to variation along one dimension while trying to ignore variation along a second dimension. The stimuli are presented in three conditions:

(1) *Single dimension* (*control*). One dimension is varied and the other is held constant.
(2) *Correlated dimensions*. Both dimensions are varied such that each value on one dimension occurs with one, and only one, of the values of the other dimension.
(3) *Orthogonal dimensions*. Both dimensions are varied such that each value on one dimension occurs with each value on the other dimension.

The subject's task is to classify the stimuli into two categories according to one designated dimension (the target dimension). If the two dimensions are perceptually *separable,* the average classification times will be the same in all three conditions. If, on the other hand, the two dimensions are perceptually *integral,* subjects will be faster in the correlated condition (*facilitation,* or redundancy gain) and slower in the orthogonal condition (*interference*) than in the single-dimension condition. The stronger the integrality, the larger the differences will be. Since several types of artifacts (such as switching attention between dimensions, contrary to instructions) can lead to facilitation in the correlated condition, interference is considered the more important result. Interference may be *symmetric* (dimension A interferes with the processing of dimension B as much as B interferes with A) or *asymmetric.*

In earlier speech perception studies, the speeded classification paradigm has been applied to investigate processing dependencies among adjacent phonetic segments (Wood & Day, 1975; Tomiak, Mullennix & Sawusch, 1987) and among phonetic features of the same segment (Eimas, Tartter, Miller & Keuthen, 1978; Soli, 1980). Most pertinent to our concerns are several studies that examined the perceptual interaction of a phonetic dimension (filter variation) with differences in F_0 (source variation). In the first of these studies, Wood (1974) used four synthesized stimuli: /bæ/ and /dæ/, with either a high (140 Hz) or a low (104 Hz) constant F_0. The target dimension was either pitch (high or low) or consonant (/b/ or /d/). Reaction times for either dimension alone were comparable. The results showed asymmetric interference: Orthogonal variation in F_0 slowed down consonant decisions, whereas orthogonal variation in consonantal information had no effect on pitch decisions. (Both dimensions, however, showed facilitation in the correlated conditions.) These results were partially replicated by Wood (1975) who employed /bæ/ and /gæ/ and omitted the correlated conditions. The F_0 variation was appropriately considered a nonlinguistic auditory dimension in these experiments, and the asymmetric interference was interpreted as supporting models of speech perception that posit an early auditory level of processing followed by a later, phonetic level. However, Pastore, Ahroon, Puleo, Crimmins, Golowner & Berger (1976) obtained similar results with analogous nonspeech stimuli consisting of a brief tone followed by a buzz, so the asymmetry may have a purely auditory origin.

Different findings emerged from a speeded classification study of the processing relationship between vowel quality and pitch (Miller, 1978). The stimuli were synthetic /bɑ/ and /bæ/, with a constant F_0 of either 104 or 140 Hz. Both dimensions interfered with each other in the orthogonal conditions to a similar degree, whereas the facilitation effects in the correlated conditions were small and nonsignificant. As in Wood's studies, the control reaction times for each dimension alone were similar. In an elaborate follow-up study, Carrell, Smith & Pisoni (1981) systematically varied the magnitudes of the vowel quality and F_0 differences, and consequently the speed of processing for each dimension. One experiment used isolated synthetic vowels, another used the same vowels preced by /b/. All stimuli had linearly falling F_0 contours, to increase their naturalness. In both experiments, Miller's finding of symmetric interference was replicated when the control reaction times for the two dimensions were equal. When the control latencies differed, however, the slower dimension showed more interference from the faster dimension than *vice versa.* (This result will be important for the interpretation of the findings of the present study.) In the second experiment, but not in the first, an asymmetry between vowel quality and pitch emerged across all discriminability conditions: the

increase in interference with the relative increase in decision times for the target dimension was steeper for vowel quality than for pitch. This asymmetry was reminiscent of that obtained by Wood, and it led Carrell *et al.* to refer once again to two-stage notions of speech perception. While isolated vowels apparently were not processed beyond the initial auditory stage, the consonantal context may have engaged the higher-level phonetic processor for the following vowel as well, thereby augmenting the pitch interference effect. An alternative explanation is that listeners detected the pitch difference during the consonantal portion at syllable onset which, according to Wood's results, does not interfere with pitch processing. A study with analogous nonspeech stimuli remains to be conducted.

The available data suggest, then, that both consonant and vowel classification times are slowed down by orthogonal variation in F_0, and also that consonantal variation does not interfere with pitch-based classification, whereas vowel quality variation does slow pitch classification but perhaps less so when a consonantal context is present. In all these studies, as noted already, F_0 was considered a nonlinguistic auditory dimension and probably was perceived as such by the subjects. Moreover, the F_0 differences were present at stimulus onset, so that minimal processing was necessary to determine a pitch category. In the present study, we were interested in whether a different pattern of mutual interference (and facilitation) would be obtained (1) when the F_0 contours are somewhat more realistic, resembling tonal contours of Mandarin Chinese, and (2) when the distinctive pitch differences are not present at stimulus onset but emerge only later, so that a significant portion of the syllable must be processed before reaching a decision. The condition of particular interest was the one that had not shown any interference in previous studies: would pitch classification be affected by consonantal variation when the F_0 contours are more natural and require more processing?

The principal purpose of the present study, however, was to compare two groups of subjects: speakers of a tone language (Mandarin Chinese) and of a nontone language (English). Conceiving of integrality as a matter of degree (cf. Garner, 1974), we hypothesized that Chinese listeners, who use tonal information together with segmental information for lexical access, will show more integrality of (i.e., more mutual interference between) segmental and tonal dimensions than do English listeners, who interpret the same tonal information as pragmatic, affective, or merely auditory variation. We also attempted to vary the lexical relevance of the tonal contours for Chinese listeners, expecting more integrality for typical Mandarin tonal contours than for atypical contours, whereas English listeners were expected to be insensitive to this difference.

We used four tasks, with the following relevant (target) and irrelevant (nontarget) dimensions: consonant/tone (C/T), vowel/tone (V/T), tone/consonant (T/C), and tone/vowel (T/V). We did not predict that effects of language background would be specific to any of these tasks. The variety of tasks was intended to lend generality to our findings, whatever they might be.

2. Methods

2.1. *Subjects*

Sixteen subjects, eight native English speakers and eight native Mandarin Chinese speakers, were recruited from the Yale University community and were paid for

their participation. The Chinese subjects were from Taiwan and were fluent in English but not in any other Chinese dialect (though they had undoubtedly been exposed to some). None of the subjects reported any history of a speech or hearing disorder. One subject of the English group was replaced because of unusually long reaction times and high error rates.

2.2. *Stimuli*

Three CV syllables, /ba/, /da/ and /bu/, were produced with four different F_0 contours by a male Mandarin Chinese speaker from Beijing, a graduate student of linguistics familiar with the purpose of our study. Two of the F_0 contours corresponded to tones 1 (high level) and 4 (high falling) of Mandarin, and each of the resulting six syllables was actually a Chinese morpheme. The other two F_0 contours (low level and low rising–falling) were uncharacteristic of Mandarin Chinese citation forms; we will refer to them as "non-Mandarin" tones, though it should be understood that Chinese listeners might nevertheless interpret them as strange renditions of familiar tones (e.g., the low level tone as tone 3). One good token of each of the 12 syllables was selected from 10 recorded repetitions. Each syllable was approximately 350 ms in duration.

In order to exclude irrelevant F_0 differences, all stimuli were LPC-coded and re-synthesized with stylized F_0 contours modeled after the originals. In particular, the two Mandarin tones (high and falling) were given the same starting frequency of 150 Hz, and the two non-Mandarin tones (low and rising–falling) both were given a starting frequency of 110 Hz. Listeners thus had to process at least part of each syllable to discriminate the two tonal alternatives. The high tone maintained the starting frequency of 150 Hz throughout, whereas the falling tone fell nonlinearly to 70 Hz. The low tone maintained its starting frequency of 110 Hz, whereas the rising–falling tone rose from 110 Hz to 150 Hz during the first 175 ms and then fell back to 110 Hz.

Eight stimulus tapes were constructed, four with Mandarin tones and four with non-Mandarin tones. The four tapes in each set represented the four tasks (C/T, V/T, T/C and T/V). Each tape included two control conditions, two correlated conditions and one orthogonal condition. The order of the conditions on each tape was fixed: control, correlated, orthogonal, correlated, control. Table I lists the stimuli employed in these conditions for the Mandarin tone tapes; the non-Mandarin tapes were analogous.

Each condition included 25 repetitions of the stimuli in random order. Thus there were 50 stimuli in each control and correlated condition, and 100 stimuli in each orthogonal condition. Altogether there were 300 stimuli on each tape. They were recorded on one channel, while an onset-synchronized mark tone was recorded on the other channel. The stimuli were separated by 1.5 s of silence. Between conditions there was a 10 s pause. One tape lasted about 13 mins.

2.3. *Procedure*

Each subject was run individually in a quiet room. The stimulus tapes were played on a Crown 800 tape recorder and were presented to the listener over TDH-39 earphones at a comfortable level. Reaction times were recorded by an Atari

TABLE I. Stimuli used in the Mandarin tone tests (H = high, F = falling)

Test	Condition				
	Control	Correlated	Orthogonal	Correlated	Control
C/T	(baH, daH)	(baH, daF)	(baH, baF, daH, daF)	(baF, daH)	(baF, daF)
V/T	(baH, buH)	(baH, buF)	(baH, baF, buH, buF)	(baF, buH)	(baF, buF)
T/C	(baH, baF)	(baH, daF)	(baH, daH, baF, daF)	(daH, baF)	(daH, daF)
T/V	(baH, baF)	(baH, buF)	(baH, buH, baF, buF)	(buH, baF)	(buH, buF)

computer whose clock was started by the mark tone and stopped by the subject's manual response.

The subject sat in front of two response keys and was asked to classify the syllables according to the target dimension (consonant, vowel, or tone) by pressing the appropriate key as rapidly as possible, trying to avoid errors. The keys were labeled /b/ and /d/ for the consonant classification tasks, /a/ and /u/ for the vowel tasks, "high" and "falling" for the Mandarin tone tasks, and "low" and "rising-falling" for the non-Mandarin tone tasks. The assignment of labels to keys was the same for all subjects.

Each subject listened to all eight test tapes, which were administered to some subjects in a single long session (with an ample break in the middle), and to other subjects in two sessions on different days. The order of the eight tapes was balanced across the eight subjects within each language group in a Latin square design.

2.4. *Analysis*

In the analysis of reaction times, the first 10 responses in each control and correlated condition and the first 20 responses in each orthogonal condition were considered practice and were discarded. This left approximately 20 responses per stimulus in each condition. All reaction times shorter than 150 ms or longer than 1 s were considered outliers and were excluded. Subsequently, correct responses were separated from errors, and the mean and standard deviation of the correct responses for each stimulus were computed. If any individual reaction time fell more than three standard deviations from the mean, it was considered an outlier and was excluded; the mean and standard deviation were recomputed until no outliers remained.

The resulting mean reaction times were subjected to a global analysis of variance, with the between-subject factor Language (Chinese, English) and the within-subject factors Task (C/T, V/T, T/C, T/V), Tone Type (Mandarin, non-Mandarin), Condition (control, correlated, orthogonal), Target Dimension (two levels), and Nontarget Dimension (two levels). Note that the design of the experiment was indeed completely factorial: each stimulus occurred in each condition (cf. Table I). Separate analyses of variance were conducted subsequently for each of the four tasks and on pairs of conditions within tasks (control *vs.* orthogonal and control *vs.* correlated), to assess interference and facilitation effects separately. The factors Target Dimension (confounded with response key) and Nontarget Dimension were involved in a number of significant effects that are of little interest here and will not

be discussed in detail.[1] Some of these effects may have been due to practice or fatigue effects that were absorbed by the "split" correlated and control conditions (cf. Table I).

3. Results

3.1. *Outliers and errors*

Table II presents the outlier and error percentages of each condition in each task, separately for Chinese and English subjects. Outliers include both correct and incorrect responses that fell either outside the absolute limits of 150–1000 ms or failed the three-standard-deviations criterion. Errors represent incorrect responses that were not outliers. While the average percentage of outliers was comparable for Chinese and English subjects (3.8% *vs.* 3.4%), the Chinese subjects made substantially fewer errors than the English subjects (1.6% *vs.* 4.4%). This was true regardless of task and condition; it seems to represent a difference in the speed-accuracy criterion adopted by the subjects (see also below). Both groups tended to make more errors in tonal than in segmental classification. Generally, error rates were not highest in the orthogonal conditions.

TABLE II. Outlier and error percentages

		Mandarin tones				Non-Mandarin tones				
		C/T	T/C	V/T	T/V	C/T	T/C	V/T	T/V	Mean
Chinese subjects										
Cont	Out	3.1	5.3	3.1	3.8	4.4	3.8	2.2	1.9	3.5
	Err	0.6	3.1	0.6	1.9	0.9	1.9	1.3	2.8	1.6
Corr	Out	4.4	4.1	3.1	4.1	2.8	5.3	4.7	3.4	4.0
	Err	1.6	1.3	0.9	1.6	0.9	1.3	1.9	0.6	1.3
Orth	Out	1.7	5.2	3.6	5.2	1.7	3.0	3.4	7.5	3.9
	Err	2.3	1.7	1.3	2.3	0.8	2.0	0.5	3.4	1.8
English subjects										
Cont	Out	2.5	5.0	4.1	5.3	3.1	3.4	1.9	5.3	3.8
	Err	4.4	5.3	2.8	3.8	5.3	9.4	7.5	3.8	5.3
Corr	Out	2.8	3.8	3.1	7.2	2.8	2.2	2.2	8.4	4.1
	Err	2.5	4.7	2.5	10.3	2.5	2.2	2.5	8.1	4.4
Orth	Out	0.6	4.1	1.7	5.5	1.3	2.5	0.9	2.3	2.4
	Err	2.3	6.6	2.5	4.5	3.0	4.8	1.6	3.1	3.6

[1] Only two effects shall be mentioned for the record, because they were so striking: the global analysis revealed a highly significant effect of Response Key (=Target Dimension) [$F(1, 14) = 20.14$, $p < 0.0006$], with right-key responses being faster than left-key responses. However, the difference was shown only by the Chinese subjects (31 ms), not by the English subjects (−1 ms); the Language by Response Key interaction was highly significant [$F(1, 14) = 22.45$, $p < 0.0004$]. This interaction did not vary significantly across tasks. All but one of our English subjects were right-handed, and so were probably all of the Chinese subjects. (We did not inquire about their handedness.) Some subjects used only one hand for responding, however. The interaction remains difficult to interpret, therefore.

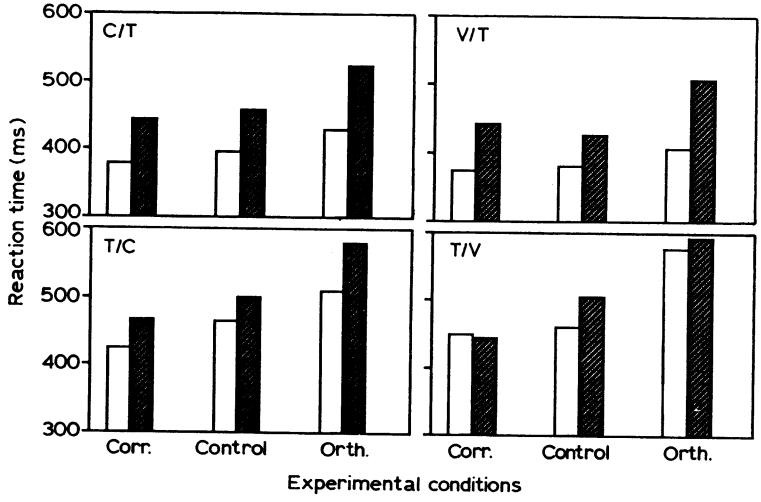

Figure 1. Average reaction times in four tasks (separate panels) as a function of Language (parameter) and Condition (abscissa). □: English; ■: Chinese.

3.2. *Reaction times*

3.2.1. *Effects of tasks and conditions*

Figure 1 shows the principal results for the reaction times (RTs) of correct responses. The four panels represent the four tasks (C/T, V/T, T/C, T/V). Each panel shows the average RTs separately for Chinese and English speakers as a function of the three experimental conditions: correlated, control and orthogonal. The control condition has been placed between the other two conditions in order to highlight *facilitation* (decrease to the left) and *interference* (increase to the right) effects.

RTs were slower in the tonal classification tasks (lower panels) than in the segmental classification tasks (upper panels); the main effect of Task was highly significant in the global analysis [$F(3, 42) = 44.28$, $p < 0.0001$]. There were also highly significant differences among the three experimental conditions [$F(2, 28) = 119.34$, $p < 0.0001$], which interacted with Task [$F(6, 84) = 13.40$, $p < 0.001$]; this justified the more detailed follow-up analyses.

Separate analyses of the four tasks showed highly significant main effects of Condition in each of them [$F(2, 28) > 29$, $p < 0.0001$]. In separate comparisons of the control and orthogonal conditions, all four tasks showed highly significant interference effects [$F(1, 14) > 35$, $p < 0.0001$]. In separate comparisons of the control and correlated conditions, three of the tasks showed significant facilitation effects, the V/T condition being the exception [C/T: $F(1, 14) = 5.63$, $p < 0.04$; V/T: $F(1, 14) = 0.45$, $p < 0.52$; T/C: $F(1, 14) = 14.44$, $p < 0.003$; T/V: $F(1, 14) = 16.33$, $p < 0.002$].

3.2.2. *Effects of language background*

Reaction times were slower for the Chinese than for the English subjects; together with the difference in error rates mentioned above, this suggests different speed-

accuracy criteria in the two subject groups. Due to large individual variability in absolute RTs, however, the difference was nonsignificant overall [$F(1, 14) = 2.15$, $p < 0.17$] and individually in all four tasks.

Slower reaction times to tonal than to segmental distinctions were shown by both language groups; however, this difference was larger for the English listeners [$F(3, 42) = 4.26$, $p < 0.02$, for the Language by Task interaction]. Moreover, the two subject groups reacted differently to the conditions in the different tasks, as shown by interactions of Language and Condition [$F(2, 28) = 3.54$, $p < 0.05$] and of Language, Task and Condition [$F(6, 84) = 3.05$, $p < 0.01$].

The question of principal interest was whether Chinese subjects showed more interference (and perhaps also more facilitation) than English subjects. In separate tests for each task, the Language by Condition interaction was significant in only one task, V/T [$F(2, 28) = 7.10$, $p < 0.04$]. However, in separate tests on the control and orthogonal conditions only (which seemed justified as planned comparisons because interference was of primary interest), significant interactions were obtained in two tasks, with a third just missing significance [C/T: $F(1, 14) = 4.27$, $p < 0.06$; V/T: $F(1, 14) = 13.28$, $p < 0.003$; T/C: $F(1, 14) = 5.37$, $p < 0.04$; T/V: $F(1, 14) = 1.14$, $p < 0.31$]. In each of the cases, the Chinese subjects showed more interference than the English subjects. Separate tests of the facilitation effects revealed a significant Language by Condition interaction only in the T/V condition [$F(1, 14) = 7.77$, $p < 0.02$], with the Chinese subjects again showing the larger effect. Note that this is the condition that did not show any difference in interference between subject groups.

These findings of larger interference (or facilitation) effects for Chinese than for English subjects are intriguing and seem to support our hypothesis. Nevertheless, they must be regarded with scepticism because of the longer RTs of the Chinese subjects. Even though the overall difference in RTs between subject groups was nonsignificant, there was reason for concern because increases in RT differences with increases in absolute RT are a common finding. Figure 2 presents scatter plots of individual subjects' average control RT *vs.* orthogonal RT, each panel representing one of the four tasks. Chinese and English subjects are represented by filled and open circles, respectively. It is evident that there was one very fast Chinese subject and one very slow English subject, hence the nonsignificant difference between the subject groups, whose RTs showed little overlap otherwise. The extent to which a data point is above the heavy diagonal line represents the magnitude of the interference effect. Regression lines were fitted to these data points; their equations are given in the figure. Slopes greater than one indicate an increase in the interference effect with absolute RT. It can be seen that the slopes were indeed greater than one in all four tasks, but not strikingly so in tasks involving vowel variation (right-hand panels). The significance of the increase was assessed by computing the correlations between the control RT and the interference effect (i.e., the difference between orthogonal and control RTs) across subjects. These correlations are given in parentheses in Fig. 2, and they were significant in only two tasks, C/T ($p < 0.05$) and T/C ($p < 0.01$). To correct for these trends, whether significant or not, the interference effects in each task were expressed as residuals from their respective regression lines, and the Language main effects were re-assessed in one-way ANOVAs. The Language effect remained significant only in

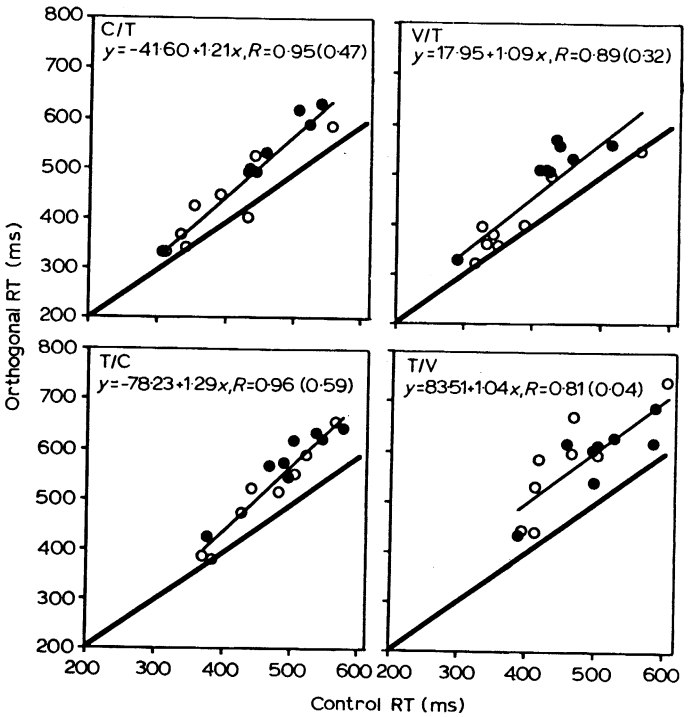

Figure 2. Scatter plots of control RT *vs.* orthogonal RT, separately for the four tasks, for individual Chinese (●) and English (○) subjects. The heavy diagonal line represents equality of the two reaction times. The thinner line is the best-fitting regression line for the whole point swarm; its equation and correlation are given in each panel. The correlation in parentheses is that between control RT and the interference effect (i.e., the difference between orthogonal and control RTs).

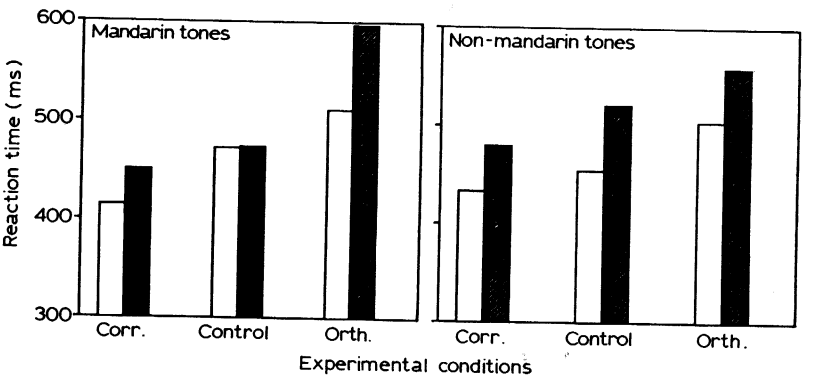

Figure 3. Average reaction times in the Tone/Consonant task as a function of Tone Type (separate panels), Language and Condition. □: English; ▨: Chinese.

the V/T condition [$F(1, 14) = 10.51$, $p < 0.006$]; that is, only in that task did the Chinese subjects show reliably more interference than the English subjects.

A similar analysis was carried out on the facilitation effects in the T/V condition. The increase in the facilitation effect with absolute control RT was nonsignificant ($r = 0.35$), and the residual Language effect—more facilitation for Chinese listeners—remained significant [$F(1, 14) = 6.29$, $p < 0.03$].

3.2.3. *Effects of tone type*

Because dimensional integrality for Chinese listeners was predicted to be stronger with Mandarin than with non-Mandarin tones, the triple interaction between Language, Condition and Tone Type was examined. In the global analysis, this interaction fell short of significance [$F(2, 28) = 2.68$, $p < 0.09$]. In the analyses of individual tasks, the interaction was significant in only one task, T/C [$F(2, 28) = 4.85$, $p < 0.02$], and specifically for the interference effect [$F(1, 14) = 6.46$, $p < 0.03$]: Chinese, but not English, subjects showed more interference with the Mandarin than with the non-Mandarin tone stimuli. This interaction is depicted in Fig. 3.

4. Discussion

The present study has addressed two issues: the general question of the perceptual integration of segmental and tonal features in speech, and the more specific question of possible differences in the strength of that integration for speakers of a tone language and a nontone language, respectively. Let us first consider the general question, with reference to speakers of English.

One surprising outcome of this study was its apparent failure to replicate the asymmetric interference between consonant place of articulation and F_0 reported previously by Wood (1974, 1975): The T/C task showed as much interference (45 ms) as the C/T task (32 ms) for English subjects, whereas Wood found average effects of about 3 and 53 ms, respectively, in similar conditions. Instead, we found asymmetric interference between vowel quality and F_0: English subjects showed substantially more interference in the T/V task (116 ms) than in the V/T task (26 ms), compared with statistically symmetric effects of 70 and 42 ms, respectively, in Miller's (1978) experiment. How are these differences to be explained?

The answer lies in the longer RTs for tonal than for segmental decisions, a difference exhibited very consistently by our subjects. In Wood's (1974, 1975) and Miller's (1978) experiments, on the other hand, the control RTs for segmental and F_0 decisions were equal. As Carrell *et al.* (1981) have shown for vowel quality *vs.* F_0, the magnitude of interference increases with the control RT for the relevant dimension. When the two dimensions differed in control RT, there was asymmetric interference, with the slower dimension showing more interference from the faster dimension than *vice versa.* Conditions with a control RT asymmetry comparable to that in our V/T and T/V tasks (panels A and B in Fig. 3 of Carrell *et al.*) indeed yielded results very similar to those of our English subjects: little or no interference of F_0 with vowel decisions, but substantial interference of vowel quality with F_0 decisions. Presumably, similar principles apply to tasks involving consonant place of articulation and F_0, so our unexpected finding of symmetric interference between these dimensions is satisfactorily explained by the longer control RTs for tones than

for consonants.[2] Our data are thus in agreement with Wood's and Miller's findings, after all.

The longer RTs for tonal than for segmental decisions were inherent in our stimuli because their F_0 contours started at the same frequency and became different only over time. We designed our stimuli in this way to force listeners to process more of the stimulus and to avoid a primitive auditory strategy of merely listening for differences in onset frequency. As far as the results for our English subjects are concerned, however, this did not seem to change the processing relationship between segmental and tonal dimensions, which remained underlyingly asymmetric for consonants and tones, but symmetric for vowels and tones.

This leads us to the more specific question concerning effects of language background. Now that we have determined that our results for English subjects are not anomalous, we are in a better position to interpret the results for the Chinese subjects. The unexpected finding that the Chinese subjects' RTs were slower overall is a problem that we tried to adjust for. Thus, the slightly larger interference effects shown by the Chinese subjects in the C/T and T/C tasks (cf. Fig. 1) should probably be disregarded as artifacts. The results in these two tasks seem rather similar for English and Chinese subjects, except for the interaction with tone type, discussed below.

Chinese and English subjects did yield different results in the V/T and T/V tasks, however: English subjects showed very different amounts of interference in the two tasks, whereas Chinese subjects showed about equal amounts (cf. Fig. 1). Since we have argued that the observed interference asymmetry between the vowel and tone dimensions for English subjects really reflects an underlying *symmetry* distorted by unequal decision times, the symmetric interference effects shown by the Chinese subjects in the face of slower control RTs for tones than for vowels (see Fig. 1) may be interpreted as an underlying *asymmetry*. In other words, the results suggest that, had the control RTs been equal, Chinese (but not English) subjects would have shown more interference in the V/T task than in the T/V task. This underlying asymmetry may reflect some of the specific linguistic experience of Chinese speakers, vowels being the carriers of lexical tones: A vowel with a different tone may seem like a different vowel to linguistically unsophisticated listeners, whereas a tone on a different vowel still sounds like the same tone.[3] In addition, Chinese subjects showed an asymmetry in facilitation effects, with no facilitation at all in the V/T task but a large effect in the T/V task (cf. Fig. 1). This suggests that they relied on the faster dimension (vowel quality) in the correlated condition of the T/V task, though it is not clear why the English subjects did not follow the same strategy.

Thus, there are some differences between Chinese and English subjects, but they are not necessarily indicative of a tighter perceptual integration of segmental and tonal dimensions in speakers of a tone language, merely of a different processing

[2] It is of interest in this connection to refer to the results of the first author, who served as a pilot subject in the study. His RTs were substantially faster than those of any regular subject in the experiment, at the cost of increased error rates. In the control conditions for segmental targets, his average RT was 249 ms (7.4% errors); in those for tonal targets, it was 324 ms (15.9% errors). The average magnitude of the interference effect was −4 ms for segmental targets and 35 ms for tonal targets. (There was an even more striking difference in facilitation effects, 4 ms *vs.* 62 ms, which probably reflects use of the correlated segmental dimension in the tonal target tasks.) These data suggest that the interference effect of tonal variation on segmental decisions can be avoided if segmental RTs are sufficiently fast.

[3] Suggested by Mary Beckman (personal communication).

relationship. Additional evidence might have been expected to emerge from interactions with the experimental variable of tone type, contrasting Mandarin and non-Mandarin tones. However, we knew this manipulation to be a weak one: considering the large interspeaker and contextual variations of tonal contours, even atypical contours can be interpreted by listeners as allophonic variants. And while such atypicality may slow down lexical access (all syllables in our study were possible meaningful words in Mandarin Chinese), we do not know to what extent automatic lexical influences may have operated in the speeded classification task; certainly, the task did not encourage lexical strategies. In any event, tone type seemed to play a significant role in only one task, T/C, where Chinese subjects showed more interference from irrelevant consonants for Mandarin than for non-Mandarin tones. They also had slower control RTs for non-Mandarin tones, which was not shown by English subjects (cf. Fig. 3). A similar, though nonsignificant, difference in control RTs was present in the T/V task. Thus there is some indication that Chinese listeners found it easier to discriminate familiar tonal contours, but at the same time they received equal or more interference from segmental dimensions, contrary to the trend of increased interference with increased control RT. Another relevant finding was that Chinese subjects were somewhat faster in responding to tonal distinctions than were English subjects, relative to their reaction times for segmental distinctions. These observations provide some additional support for our hypothesis of differences in perceptual processing as a function of language background.

Clearly, our study is not the last word on this issue. It used only a single experimental paradigm, a limited set of simple stimuli, and relatively small groups of subjects. The Chinese subjects, moreover, were not monolingual, having had extensive experience with English. Ideally, one would like to conduct this type of study with Chinese speakers who have not been exposed to a nontone language, but they are difficult to find outside China. Further research will be needed to determine whether Chao's (1980) observation, quoted above, truly referred to perception, or whether it merely paraphrased the truism that different tones makes different words in a tone language.

This research was supported by NICHD Grant HD01994 to Haskins Laboratories. We are grateful to Yi Xu for running the subjects and processing the data. When this research was begun, the second author was affiliated with Haskins Laboratories and the University of Connecticut. Address correspondence to Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695, U.S.A. or to Hwei-Bing Lin, Department of Speech and Hearing Sciences, The Graduate Center, CUNY, 33 W. 42 Street, New York, NY 10036, U.S.A.

References

Basso, A., Casati, G. & Vignolo, L. A. (1977) Phonemic identification defect in aphasia, *Cortex*, **13**, 85–95.

Blumstein, S. & Cooper, W. E. (1974) Hemispheric processing of intonation contours, *Cortex*, **10**, 154–168.

Blumstein, S. E., Baker, E. & Goodglass, H. (1977) Phonological factors in auditory comprehension in aphasia, *Neuropsychologia*, **15**, 19–30.

Carrell, T. D., Smith, L. B. & Pisoni, D. B. (1981) Some perceptual dependencies in speeded classification of vowel color and pitch, *Perception & Psychophysics*, **29**, 1–10.

Chao, Y. R. (1980) Chinese tones and English stress. In *The Melody of Language* (L. R. Waugh & C. H. Schooneveld, editors), pp. 41–44. Baltimore, MD: University Park Press.

Divenyi, P. L. & Robinson, A. J. (1989) Nonlinguistic auditory capabilities in aphasia, *Brain and Language*, **37**, 290–326.
Edmondson, J. A., Chan, J.-L., Seibert, G. B. & Ross, E. D. (1987) The effect of right-brain damage on acoustical measures of affective prosody in Taiwanese patients, *Journal of Phonetics*, **15**, 219–233.
Eimas, P. D., Tartter, V. C., Miller, J. L. & Keuthen, N. J. (1978) Asymmetric dependencies in processing phonetic features, *Perception & Psychophysics*, **23**, 12–20.
Fant, G. (1960) *Acoustic Theory of Speech Production*. The Hague: Mouton.
Gandour, J. & Dardanaranda, R. (1983) Identification of tonal contrasts in Thai aphasic patients, *Brain and Language*, **18**, 98–114.
Gandour, J., Petty, S. H. & Dardaranada, R. (1988) Perception and production of tone in aphasia, *Brain and Language*, **35**, 201–240.
Garner, W. R. (1970) The stimulus in information processing, *American Psychologist*, **25**, 350–358.
Garner, W. R. (1974) *The Processing of Information and Structure*. Potomac, MD: Erlbaum.
Garner, W. R. & Felfoldy, G. L. (1970) Integrality of stimulus dimensions in various types of information processing, *Cognitive Psychology*, **1**, 225–241.
Goodglass, H. & Calderon, M. (1977) Parallel processing of verbal and musical stimuli in right and left hemispheres, *Neuropsychologia*, **15**, 397–407.
Halperin, Y., Nachshon, I. & Carmon, A. (1973) Shift of ear superiority in dichotic listening to temporally patterned nonverbal stimuli, *Journal of the Acoustical Society of America*, **53**, 46–50.
Hartje, W., Willmes, K. & Weniger, D. (1985) Is there parallel and independent hemispheric processing of intonational and phonetic components of dichotic speech stimuli? *Brain and Language*, **24**, 83–99.
Kimura, D. (1967) Functional asymmetry of the brain in dichotic listening, *Cortex*, **3**, 163–178.
Ley, R. G. & Bryden, M. P. (1982) A dissociation of right and left hemispheric effects for recognizing emotional tone and verbal content, *Brain and Cognition*, **1**, 3–9.
Miller, J. L. (1978) Interactions in processing segmental and suprasegmental features of speech, *Perception & Psychophysics*, **24**, 175–180.
Naeser, M. A. & Chan, S. W.-C. (1980) Case study of a Chinese aphasic with the Boston Diagnostic Aphasia Exam, *Neuropsychologia*, **18**, 389–410.
Packard, J. L. (1986) Tone production deficits in nonfluent aphasic Chinese speech, *Brain and Language*, **29**, 212–223.
Pastore, R. E., Ahroon, W. A., Puleo, J. S., Crimmins, D. B., Golowner, L. & Berger, R. S. (1976) Processing interaction between two dimensions of nonphonetic auditory signals, *Journal of Experimental Psychology: Human Perception and Performance*, **2**, 267–276.
Shipley-Brown, F., Dingwall, W. O., Berlin, C. I., Yeni-Komshian, G. & Gordon-Salant, S. (1988) Hemispheric processing of affective and linguistic intonation contours in normal subjects, *Brain and Language*, **33**, 16–26.
Soli, S. D. (1980) Some effects of acoustic attributes of speech on the processing of phonetic feature information, *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 622–638.
Square-Storer, P., Darley, F. L. & Sommers, R. K. (1988) Nonspeech and speech processing skills in patients with aphasia and apraxia of speech, *Brain and Language*, **33**, 65–85.
Studdert-Kennedy, M. & Shankweiler, D. (1970) Hemispheric specialization for speech perception, *Journal of the Acoustical Society of America*, **48**, 579–594.
Tomiak, G. R., Mullennix, J. W. & Sawusch, J. R. (1987) Integral processing of phonemes: Evidence for a phonetic mode of perception, *Journal of the Acoustical Society of America*, **81**, 755–764.
Van Lancker, D. & Fromkin, V. A. (1973) Hemispheric specialization for pitch and "tone": Evidence from Thai, *Journal of Phonetics*, **1**, 101–109.
Van Lancker, D. & Fromkin, V. A. (1978) Cerebral dominance for pitch contrasts in tone language speakers and in musically untrained and trained English speakers, *Journal of Phonetics*, **6**, 19–23.
Wood, C. C. (1974) Parallel processing of auditory and phonetic information in speech discrimination, *Perception & Psychophysics*, **15**, 501–508.
Wood, C. C. (1975) Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses, *Journal of Experimental Psychology: Human Perception and Performance*, **104**, 3–20.
Wood, C. C. & Day, R. S. (1975) Failure of selective attention to phonetic segments in consonant-vowel syllables, *Perception & Psychophysics*, **17**, 346–350.
Zurif, E. & Mendelsohn, M. (1972) Hemispheric specialization for the perception of speech sounds: The influence of intonation and structure, *Perception & Psychophysics*, **11**, 329–332.