Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese

LISA LEE and HOWARD C. NUSBAUM
University of Chicago, Chicago, Illinois

The processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese were investigated in a speeded classification task. Since in Chinese, unlike in English, tones convey lexically meaningful information, native speakers of these languages may process combinations of segmental and suprasegmental information differently. Subjects heard consonant-vowel syllables varying on a consonantal (segmental) dimension and either a Mandarin Chinese or constant-pitch (non-Mandarin) suprasegmental dimension. The English listeners showed mutual integrality with the Mandarin Chinese stimuli, but not the constant-pitch stimuli. The native Chinese listeners processed these dimensions with mutual integrality for both the Mandarin Chinese and the constant-pitch stimuli. These results were interpreted in terms of the linguistic function and the structure of suprasegmental information in Chinese and English. The results suggest that the way listeners perceive speech depends on the interaction between the structure of the signal and the processing strategies of the listener.

In recognizing spoken words, listeners interpret information from the patterns of speech using a variety of sources of linguistic knowledge. Knowledge of the semantic, syntactic, and phonological structure of language, for example, provides much constraint for word recognition (e.g., Newell, 1975). However, even when considering the pattern structure of speech alone, different kinds of information contribute to recognition. Listeners recognize speech using both *segmental* information, which concerns the consonants and vowels in speech, and *suprasegmental* information, which concerns acoustic properties that extend over more than one segment, such as intonation contours or stress patterns. In all languages, segmental distinctions are used to convey differences between words; however, in some languages, suprasegmental information serves this function as well. In tone languages such as Mandarin Chinese, two different words may have exactly the same pattern of consonants and vowels and differ only in their pattern of intonation. Every word in Mandarin Chinese has one of four tones; placing a different tonal contour on the same segmental sequence can change word meaning. For example, the word *da* may mean *dozen*, *hit*, or *big*, depending on the tone applied to it.

In contrast to Chinese, in languages like English, suprasegmentals have a much more limited role in distinguishing words. For example, stress differences sig-

nal the noun-verb distinction in words such as *rebel*, in which primary stress falls on the first syllable for a noun and on the second syllable for a verb (see Chomsky & Halle, 1968). However, beyond this relatively limited lexical function, intonation contours in English generally convey syntactic, pragmatic, and affective information (see Bolinger, 1989). The way native Chinese and English listeners represent words may reflect the different lexical function of suprasegmentals. For example, native Chinese listeners may incorporate both segmental and suprasegmental information in their lexical representations, whereas native English listeners may represent primarily segmental information. As a consequence of the differences in lexical relevance of segmental and suprasegmental information to native Chinese and English listeners, they may show different patterns of perceptual interactions between these two types of information. That is, native English listeners may process segmental and suprasegmental dimensions as different kinds of information on the basis of their different phonemic status, whereas native Chinese listeners may process these dimensions similarly on the basis of their shared phonemic status. Furthermore, since in Mandarin Chinese both suprasegmental and segmental information play a phonemic role in recognizing words, perhaps these dimensions are processed by native listeners of Chinese in the same way that segmental (i.e., phonemic) dimensions are processed by English listeners.

An experimental paradigm that reveals the nature of the interactions between different sources of information is Garner's (1970, 1974) speeded classification task. In this paradigm, subjects hear stimuli that can vary along two dimensions and classify them according to their values

This research was supported in part by National Institute of Deafness and Other Communicative Disorders, DC 00601. We thank Xiaolei Wang for her advice and assistance with the Mandarin Chinese materials. We also thank Jenny DeGroot and Anne Henly for helpful comments on an earlier draft of this manuscript. Address correspondence and reprint requests to H. C. Nusbaum, Department of Psychology, University of Chicago, 5848 S. University Ave., Chicago, IL 60637.

on a target dimension. If two dimensions are processed integrally—that is, if processing of one dimension entails processing of the other as well—listeners will have difficulty selectively attending to only one dimension. Wood and Day (1975) have presented evidence that native English listeners process segmental dimensions integrally. They presented listeners with stimuli varying along two segmental dimensions, consonant identity (/b/ vs. /d/) and vowel identity (/a/ vs. /ae/), in CV syllables. In a *control* condition, subjects were presented with repetitions of two stimuli that varied along only one dimension, the target dimension, while the other dimension was held constant. Listeners judged each stimulus according to its value on the target dimension when there was no variation in the nontarget dimension. For example, in one block of trials listeners judged target consonant identity (/b/ vs. /d/) in repetitions of the syllables /ba/ and /da/. In an *orthogonal* condition, subjects were presented with stimuli that varied along both dimensions. They classified each stimulus on the target dimension in the context of irrelevant variation in the nontarget dimension.¹ For example, listeners judged target consonant identity (/b/ vs. /d/) in presentations of /ba/, /bae/, /da/, and /dae/. If the nontarget dimension (in this case, /a/ vs. /ae/) must be processed in conjunction with the target dimension, then irrelevant variation in this dimension will increase processing time. Thus, integrality between dimensions is indicated by longer response times (RTs) in the orthogonal condition than in the control condition. If instead two dimensions are separable, RTs in the control and orthogonal conditions will not be statistically different. Wood and Day (1975) found that whether attending to a consonant or vowel target dimension, listeners were slowed in their responses by variation in the nontarget dimension. These results demonstrate that segmental dimensions are processed integrally by native English listeners.

Further evidence shows that this integrality is a function not of acoustic features of the stimuli but of the listener's perceptual interpretation of the dimensions. Tomiak, Mullennix, and Sawusch (1987) demonstrated that when presented with noise-tone analogs of fricative-vowel syllables and told that these syllables are nonspeech, listeners do not process them integrally. However, when told that the noise-tone analogs are speech, listeners process them integrally. These results suggest that if segmental and suprasegmental dimensions are interpreted similarly according to linguistic function, they may show interactions in processing as well. In contrast, if these dimensions have different linguistic functions, they may instead be processed separably.

Investigations of the integrality between segmental and suprasegmental dimensions, however, have yielded findings of both integrality and separability between these dimensions. Wood (1974, 1975) examined the interactions between a phonetic dimension, the place contrast /b/ versus /g/ in the context of the vowel /ae/, and a suprasegmental dimension, a low-level pitch (104 Hz) versus a high-level pitch (140 Hz). The results showed asym-

metric integrality between these dimensions for native English listeners. When the listeners were judging pitch, variation in place did not slow their RTs. However, in contrast to the prediction that the dimensions of pitch and phoneme should be separable for English listeners, Wood (1974, 1975) found that when English listeners were judging place of articulation, variation in pitch did slow their RTs. Wood interpreted these results as evidence for two levels of processing, in which pitch information is processed at an auditory level prior to phonetic processing. According to this account, if pitch information is processed only at this initial auditory level, then it will not be affected by phonemic variation, since perception of this variation occurs later when processed at a subsequent phonetic level. Conversely, the processing of phonemes will be affected by earlier processing of auditory (including pitch) information, since it is the output of this earlier processing that is then processed at the phonetic stage.

The difference in phonological status between segmental and suprasegmental information is clear in English—consonants and vowels are phonemic and suprasegmentals are not. If the difference between the processing stages Wood (1974, 1975) proposes is based on linguistic function (e.g., segmental vs. suprasegmental), then vowels should produce the same processing interactions as consonants. However, if the distinction is governed mainly by auditory characteristics of the stimulus, then vowels should function like suprasegmentals. Although the results of Tomiak et al. (1987) suggest that perceptual function, rather than acoustic characteristics, of dimensions should govern processing interactions, an investigation of the processing interactions between vowel and pitch (Miller, 1978) supports an acoustic basis for these interactions. Vowels and consonants are both phonemic, but they differ in their acoustic characteristics. Rapid changes in amplitude and fundamental frequency characterize the acoustic cues to consonant identity (Delattre, Liberman, & Cooper, 1955), whereas more steady-state acoustic information characterizes the cues to vowel identity (Fry, Abramson, Eimas, & Liberman, 1962). Miller found that vowels do not produce the same patterns of processing as consonants. Rather, her findings show that native English listeners process vowels and pitch with mutual and symmetric integrality. These findings support the suggestion that the ways in which dimensions interact in processing depend on the acoustic characteristics of the information being processed.

If differing acoustic characteristics primarily govern the nature of dimensional interactions, then native listeners of different languages should show no difference in their processing of segmental and suprasegmental information. That is, despite language-specific differences in the function of suprasegmentals in languages such as Chinese and English, native listeners of these languages should show similar patterns of processing. Repp and Lin (1990) examined whether differences in the phonological function of suprasegmentals in Chinese and English result in different strategies for processing this information in native

listeners of these languages. They tested native English and Chinese listeners on their perception of segmental (consonant, vowel) and suprasegmental (Mandarin tones, non-Mandarin tones) information in the speeded classification task (Garner, 1970, 1974). Analyses of the effect of native language on integrality showed that these listeners performed quite similarly on the classification tasks. Both groups showed integrality between segmental and suprasegmental sources of information for all dimensional combinations and classification judgments. However, the Chinese and English listeners did differ in the amount of integrality they displayed between dimensions. The Chinese listeners appeared to show greater integrality for one of four tasks (vowel judgments in context of varying tone) and greater integrality for the Mandarin than the non-Mandarin tones in one of four tasks (tone judgments in context of varying consonants). The overall similarity between Chinese and English listeners suggests that an explanation based on the lexical (i.e., tonal) function of the suprasegmental information does not specify the characteristics of dimensional interactions.

The similar performance of Repp and Lin's native Chinese and English listeners supports the notion that acoustic characteristics of the stimulus govern perceptual interactions. However, how do we reconcile the differences in the patterns of perceptual interactions found by Wood (1974, 1975) and by Repp and Lin (1990)? Wood's data show that English listeners process suprasegmental (pitch) dimensions independently of segmental (consonant) dimensions when focusing on suprasegmental judgments. However, Repp and Lin's English listeners showed mutual and symmetric integrality between these dimensions, a result that conflicts with Wood's levels-of-processing explanation. Repp and Lin note that differences between their findings and those of Wood may be due to differences in the relative discriminability of dimensions. Patterns of perceptual integrality may change as the relative discriminability of dimensions is varied (e.g., Carrell, Smith, & Pisoni, 1981; but see Eimas, Tartter, Miller, & Keuthen, 1978). In Repp and Lin's study, discriminability varied; subjects showed longer control RTs (lower discriminability) for tonal, as opposed to segmental, dimensions. Had the discriminability of the suprasegmental dimension in their study been increased, consonant and pitch may have shown asymmetric integrality, as in the Wood studies.

Although a discriminability explanation may account for the differences between Repp and Lin's (1990) and Wood's (1974, 1975) results, another explanation is also possible. Just as a difference in the acoustic characteristics of consonant and vowel segments may affect integrality, so may differences in types of suprasegmental information. Whereas the suprasegmental dimension for the Wood studies consisted only of level pitches, Repp and Lin used pairs of dynamic pitches or combinations of dynamic and static (level) pitches. A possible explanation for the differences in the patterns of results in these studies is that the particular suprasegmental dimensions incorporated in

each study are processed differently. Native Chinese listeners' processing of static pitches and segmentals will be relevant to assessing processing of different types of suprasegmentals by listeners from different language backgrounds.

The goal of the present study was to investigate further the processing interactions between different kinds of suprasegmental and segmental information and to examine how processing of these dimensions may differ in listeners from different native language backgrounds. The processing interactions between a segmental dimension and two kinds of suprasegmental dimensions were examined. The segmental dimension consisted of a consonantal contrast between /ba/ and /da/. These syllables were paired with two different types of suprasegmental dimensions, Mandarin tones and (non-Mandarin) constant pitches. The Mandarin tones were two dynamic contours corresponding to Tones 3 and 4 in Mandarin. These particular tones were chosen to discover whether Repp and Lin's results could be replicated with a different set of Mandarin tones. The constant pitches were a low pitch and a high pitch, chosen to match the suprasegmental dimension of the Wood (1974, 1975) stimuli. Thus, two sets of stimuli, four Mandarin syllables and four constant-pitch syllables, were presented to subjects for speeded classification in the Garner (1970, 1974) paradigm. Two groups of subjects, native Mandarin Chinese and native English listeners, participated in the experiment.

The present study was carried out to clarify the combined roles of stimulus characteristics and characteristics of the listener's linguistic experience on the processing of segmental and suprasegmental sources of information. For the dimensions on which both Chinese and English listeners have been tested so far (Repp & Lin, 1990), they show similar patterns in processing of segmental and suprasegmental sources of information. The present study extends the comparison of Chinese and English listeners' processing to different suprasegmental dimensions. If the way listeners process segmentals and suprasegmentals depends on the particular characteristics of the stimulus dimensions, regardless of their linguistic relevance, then the native Chinese and English listeners should continue to resemble each other in their patterns of perceptual integrality. However, if native language influences processing strategies, the patterns of integrality that the native Chinese and English listeners display could be different for the different pairings of segmental and suprasegmental dimensions. Because of the function of tone in Chinese, native Chinese listeners may again show integrality for all pairings of segmental and suprasegmental dimensions regardless of lexical function, including the constant-pitch condition for which native English listeners show asymmetric integrality (Wood, 1974, 1975). This pattern of results for the Chinese listeners would be consistent with Repp and Lin's findings with native Chinese listeners (Repp & Lin, 1990). An analogous prediction for the native English listeners would be that, because of the non-lexical function of tone in English, native English listeners

may not show integral processing for all types of segmental and suprasegmental information. However, given the differences in the results reported by Wood (1974, 1975) and Repp and Lin (1990), no single prediction can be made about the effects of linguistic experience on the integrality of these stimulus dimensions for English listeners.

METHOD

Subjects

Seventeen subjects between the ages of 18 and 41 participated in the experiment. All were students or staff at the University of Chicago or residents of the neighborhood. Eight of these, 6 males and 2 females, were native speakers of Mandarin Chinese who came to the university from the People's Republic of China. Although some of the native Mandarin speakers had been exposed to other dialects, none were fluent in those dialects. Nine subjects, 5 males and 4 females, were native speakers of English, with no experience speaking Mandarin. None of the subjects reported speech or hearing disorders. Each participated in two 1-h sessions and was paid \$10 after completing the second session.

Stimuli

The stimuli were eight syllables generated on the Klatt speech synthesizer (Klatt, 1980a). For the constant-pitch stimuli, four syllables were created with the same suprasegmental dimension as the stimuli of Wood (1974, 1975). In this stimulus set, the four syllables consisted of /ba/ and /da/, each produced at a low fundamental frequency (F0) and at a high F0. The syllables /ba/ and /da/ were chosen because they yield real lexical items in Chinese. In the Mandarin stimulus set, the four syllables were /ba/ and /da/, each produced with a low-rising tone (third tone) and a falling tone (fourth tone). The syllable /ba/ with the third tone refers to a word that functions as a syntactic marker and also means *to hold with the hand*. The syllable /da/ with the third tone means *to hit*. The syllables /ba/ and /da/ with the fourth tone mean *father* and *big*, respectively.

The synthesis parameters for the consonant and vowel of all four /ba/ syllables were identical: These syllables differed only in their F0 contours. Similarly, all four /da/ syllables were identical except for their F0 contours. All stimuli were 300 msec in duration. The amplitude of each syllable was ramped up from 5 to 60 dB in the first 20 msec of the stimulus, and remained at 60 dB for the duration of the syllable. For the /ba/ syllables, the starting and steady-state frequencies for the first three formants (F1, F2, and F3) were 280 and 700 Hz, 1113 and 1220 Hz, and 2173 and 2600 Hz, respectively. The formant transition periods were 40 msec for F1, 55 msec for F2, and 65 msec for F3. For the /da/ syllables, the starting and steady-state frequencies were 200 and 700 Hz for F1, and 1520 and 1220 Hz for F2. The formant transition periods were 65 msec for F1 and 90 msec for F2. F3 was held constant at 2600 Hz.

In the constant-pitch stimulus set, F0 was set at 104 Hz for the low-pitch syllables and at 140 Hz for the high-pitch syllables. To create the contours for the Mandarin stimulus set, a native speaker of Mandarin was asked to produce tokens of /ba/ and /da/ with the third and fourth tones. The F0 contours of these utterances were examined, and stylized versions of these contours were added to the synthetic /ba/ and /da/ syllables. In the syllables with a third tone, F0 at the beginning of the syllable was 137 Hz, dropping to 84 Hz at 165 msec, and ending at 102 Hz. In the syllables with a fourth tone, F0 started at 165 Hz and fell linearly to 95 Hz by the end of the syllable. These tonal contours are illustrated in Figure 1.

To confirm that the Mandarin stimuli are heard as Mandarin and that the constant-pitch stimuli are not, two native speakers of Mandarin, neither of whom participated as a subject in the speeded clas-

Figure 1. Stylized tonal contours for the Mandarin syllables with third tone (top panel) and fourth tone (bottom panel).

sification task, were asked to judge the quality of the stimuli. In separate blocks, they heard five repetitions in random order of the constant-pitch and then the Mandarin stimuli and were asked to write down, in any language they desired, what the stimuli sounded like to them. The blocks were then repeated, with the order of the stimulus sets reversed, and the listeners were asked to interpret each stimulus as if it were a Mandarin syllable. The results showed that these listeners had little difficulty identifying the segmental dimension as /ba/ or /da/; segmental accuracy averaged 97% across listeners and blocks. When judging the constant-pitch stimuli in any language, both listeners transcribed the syllables in the Roman alphabet, with no tone markings. When asked to interpret these syllables as Mandarin, on 95% of trials (i.e., on 9 of 10 trials for one listener, and 10 of 10 for the other), listeners labeled /da/-104 Hz and /da/-140 Hz identically (as /da/ with a first tone, which may mean *to lay across*, *lift*, *to take a means of transportation*, or *add*) despite the suprasegmental difference. With /ba/, one listener distinguished the difference in pitch on all trials, interpreting the high pitch as Tone 1 (meaning *eight*) and the low pitch as Tone 3 (the syntactic marker or *to hold*). This listener noted that the low-level pitch was a poor example of the third tone. The other listener labeled both as Tone 1 on all trials. In contrast, when labeling the Mandarin stimuli under the instructions to do so in any language, one listener transcribed them as Chinese characters, while the other listener transcribed them as pinyin (an alphabetized transcription including the appropriate tone markings). Both listeners heard the suprasegmental contrast, third versus fourth tone, as we intended

and with no errors. When then asked to transcribe the Mandarin syllables as Mandarin, they again interpreted the syllables accurately and with appropriate lexical interpretations.

During subject testing, stimuli were converted in real time to analog form under computer control at 10 kHz with a 12-bit D/A converter. The speech was lowpass filtered at 4.6 kHz and presented at about 74 dB SPL over Sennheiser HD-430 headphones.

Procedure

The subjects participated in two 1-h sessions conducted on separate days within a 1-week period. In one session, the subjects performed the speeded classification task with the constant-pitch stimulus set; in the other session, they performed the same task with the Mandarin stimulus set. The order in which the subjects completed these sessions was counterbalanced. A session consisted of eight blocks: A set of four consecutive blocks of trials was presented for each of two judgment tasks (segmental and suprasegmental). Half the subjects performed segmental judgments first, and half performed suprasegmental judgments first. The first block in each set of four was always a practice block in which subjects heard and responded to the syllables in the stimulus set and received feedback on their responses. The practice block was followed by three test blocks, two control and one orthogonal. Although the two control blocks were always presented consecutively and in the same order, the test blocks were counterbalanced such that half the subjects always received the paired control blocks first and half received the orthogonal first.

For each of the two 1-h testing sessions stimuli were grouped for two pairs of control blocks and two orthogonal blocks. In each pair of control blocks, the subjects heard two stimuli in which the values on one dimension varied and the values on the other dimension were fixed. In each orthogonal block, the subjects heard four stimuli in which values on both dimensions varied. For example, in half of the constant-pitch testing session, the subjects judged segment identity (/b/ vs. /d/). Two members of the constant-pitch stimulus set, /ba/-104 Hz and /da/-104 Hz, comprised one of a pair of control blocks and the other two members, /ba/-140 Hz and /da/-140 Hz, comprised the other control block. The entire set of four stimuli comprised the orthogonal block. The blocks were arranged such that the same stimuli were included in both a control block and its corresponding orthogonal block. Thus, each stimulus served as its own control across conditions. For a complete listing of the stimuli used for each judgment task for each condition and test session, see Tables 1 and 2.

In the practice blocks that preceded each set of control and orthogonal blocks, each member of the stimulus set was presented in random order a total of twice each. In the control and orthogonal blocks, stimuli were presented 20 times each in random order. Thus, each control block consisted of 40 trials and each orthogonal block consisted of 80 trials. Response keys were labeled as *b* and *d*, *high*

Table 2
Stimuli for Each Condition in the Constant-Pitch Session

Dimension	Condition	
	Control	Orthogonal
Consonant	/ba/-low	/ba/-low
	/da/-low	/da/-low
	or	/ba/-high
	/ba/-high	/da/-high
	/da/-high	
Pitch	/ba/-low	/ba/-low
	/ba/-high	/da/-low
	or	/ba/-high
	/da/-low	/da/-high
	/da/-high	

and *low*, or *3rd* and *4th*, for the segmental, suprasegmental-pitch, and suprasegmental-tone judgment tasks, respectively. The assignment of responses to hands was counterbalanced across subjects.

All instructions were recorded in advance and played to subjects on cassette tape. The English listeners received instructions in English, and the Chinese listeners received instructions in Mandarin Chinese. All subjects were instructed that they would hear repetitions of several syllables and that their task would be, depending on the block, to decide which consonant or tone/pitch they heard and to press the appropriate response key as quickly as possible. The segmental and suprasegmental dimensions of the syllables were described and labeled for the subjects. The constant-pitch stimulus set was described as a set of syllables spoken at low and high pitch. The Mandarin stimuli were described as real lexical items in Chinese, and subjects were told their meanings. In addition to the procedures followed for both groups of subjects, the Chinese subjects were shown the Chinese characters that corresponded to each of the syllables in the Mandarin stimulus set.

Experimental sessions were conducted individually. At the beginning of each trial, the subjects saw the signal READY on a computer screen. Following the ready signal, the response choices (*b* or *d*, *high* or *low*, *3rd* or *4th*) appeared on the screen. Next, the subjects heard a single syllable through the headphones, and they responded by pressing one of the designated keys on a computer-controlled keyboard. For the Chinese subjects, the response choices that appeared on the screen during presentation of the Mandarin stimuli were supplemented by pinyin transcriptions of the stimuli. The subjects were told how to interpret the pinyin; none had difficulty understanding this writing system.

RESULTS

The subjects performed the speeded classification task quite accurately. The native English group averaged 98.0% correct classification across conditions, and the Chinese group averaged 98.6% correct. Although both groups were similarly accurate in responding to stimuli [$t(15) = -.81$, n.s.], the Chinese subjects were 175 msec slower overall (averaged across all trials) than the English listeners [$t(15) = -2.89$, $p < .01$]. Repp and Lin (1990) reported that their Chinese listeners also had longer RTs than did their English listeners, and since their Chinese listeners were also substantially more accurate, they attributed the pattern of RTs and accuracy data to a speed-accuracy tradeoff. Since both groups in the present study were highly and comparably accurate, there is no evidence for a speed-accuracy tradeoff, although

Table 1
Stimuli for Each Condition in the Mandarin Session

Dimension	Condition	
	Control	Orthogonal
Consonant	/ba/-3rd tone	/ba/-3rd tone
	/da/-3rd tone	/da/-3rd tone
	or	/ba/-4th tone
	/ba/-4th tone	/da/-4th tone
	/da/-4th tone	
Tone	/ba/-3rd tone	/ba/-3rd tone
	/ba/-4th tone	/da/-3rd tone
	or	/ba/-4th tone
	/da/-3rd tone	/da/-4th tone
	/da/-4th tone	

the high level of accuracy could mask any such differences that exist. However, since the patterns of perceptual integrality within groups of listeners with the same language background are of main interest in the present study, the difference in overall speed of response between the native Chinese and English subjects is not problematic.

In scoring the RT data for each subject, trials for which the RT was more than 2.5 standard deviations above the subject's mean RT for the block were discarded, and new block means were computed over the remaining trials. The mean percentage of discarded trials was 2.3% for the English listeners and 2.6% for the Mandarin listeners. Although for each judgment task the control condition was presented in two blocks, one for each level of the irrelevant dimension that was held constant, the means for these paired control blocks were averaged together for comparison with performance in the orthogonal block.

In examining perceptual integrality, patterns of RTs in the control versus orthogonal conditions were compared. To evaluate how the native language background of the listener and the acoustic characteristics of the stimuli influence perceptual integrality, eight planned comparisons were carried out. These planned comparisons assess dimensional integrality for each combination of language group (Chinese, English), stimulus set (Mandarin, constant-pitch), and judgment condition (segmental, suprasegmental).

The RTs of the native English listeners in each condition for each judgment task are shown in Table 3. Any difference between the Mandarin and constant-pitch suprasegmentals is of particular importance in determining the effects of type of suprasegmental on the integrality of stimulus dimensions. The planned comparisons showed that when making segmental judgments, English listeners are slower in the orthogonal than in the control condition for both the Mandarin stimuli [$F(1,8) = 40.78$, $p < .01$] and the constant-pitch stimuli [$F(1,8) = 22.79$, $p < .01$].² Thus, English listeners are affected by irrelevant variation in the suprasegmental dimension when they are attending to the segmental dimension for both sets of

stimuli. This finding is consistent with the results reported previously by Wood (1974, 1975), Repp and Lin (1990), and Miller (1978). For the suprasegmental judgments, a different pattern of results was obtained. As with the segmental judgments, the English listeners are significantly slower in the orthogonal condition for the Mandarin stimuli [$F(1,8) = 6.85$, $p < .05$]. However, this is *not* the case for the constant-pitch stimuli [$F(1,8) = 2.85$, $p > .12$]. That is, irrelevant segmental variation affects English listeners when they are attending to dynamic tonal contours but not level pitches. Thus, for the segmental and suprasegmental judgments of the constant-pitch stimuli, our results replicate Wood's finding. Likewise, for the segmental and suprasegmental judgments of the Mandarin stimuli, the performance of the English listeners is consistent with Repp and Lin's (1990) findings of mutual and symmetric integrality between segmentals and both Mandarin and non-Mandarin suprasegmentals. The lack of integrality for suprasegmental judgments of constant pitches, however, contrasts with Repp and Lin's findings of integrality with other suprasegmental dimensions.³

The RTs of the native Chinese listeners for each condition and judgment are also listed in Table 3. The planned comparisons for these subjects indicate that in making segmental judgments, Chinese listeners are slowed by orthogonal variation in suprasegmental context for both the Mandarin stimuli [$F(1,7) = 7.40$, $p < .05$] and the constant-pitch stimuli [$F(1,7) = 5.47$, $p < .05$].⁴ Similarly, when making suprasegmental judgments, they are slowed by orthogonal variation in segmental context for both types of stimuli [$F(1,7) = 8.31$, $p < .05$, for Mandarin; $F(1,7) = 18.09$, $p < .01$, for constant-pitch]. These planned comparisons thus show that for Chinese listeners segmental and suprasegmental sources of information are perceived integrally for both the constant-pitch and the Mandarin stimuli. This finding of integrality when listeners are making suprasegmental judgments contrasts with Wood's (1974, 1975) results, but the finding of mutual orthogonal interference for segmental and suprasegmental judgments replicates Repp and Lin's (1990) findings.⁵

Since the relative discriminability of dimensions, as measured by differences in control RTs, may affect interpretations concerning perceptual integrality (e.g., Carrell et al., 1981; but see Eimas et al., 1978), *t* tests were conducted to determine whether discriminability differed between dimensions for the relevant comparisons. These *t* tests indicated that, for the native Chinese listeners, the relative discriminability of the segmental and suprasegmental dimensions of the stimuli did not differ for the Mandarin stimulus set [$t(7) = -.381$, $p > .35$] or for the constant-pitch stimulus set [$t(7) = -.182$, $p > .43$]. For the native English listeners, relative discriminability did not differ for the constant-pitch stimuli [$t(8) = -.842$, $p > .21$], but it did differ for the Mandarin stimuli [$t(8) = -2.80$, $p < .05$]. This finding of a difference in discriminability of dimensions for the Mandarin syllables indicates that the consonant dimension was more discriminable than the tone dimension for the native English

Table 3
Mean Response Times in Milliseconds for
Each Language Group and Stimulus Set

	Control	Orthogonal
Mandarin Stimulus Set		
English		
Consonant	474	529
Tone	540	609
Chinese		
Consonant	683	774
Tone	711	813
Constant-Pitch Stimulus Set		
English		
Consonant	492	558
Pitch	519	551
Chinese		
Consonant	619	713
Pitch	623	737

listeners. The Mandarin tones in these stimuli do differ by 28 Hz in frequency at onset, and so could be immediately discriminated on that basis by listeners. However, the initial direction of frequency change for both tones is in a falling direction (see Figure 1). This characteristic of the stimuli might have made the tone more difficult to discriminate than the constant pitches, which differ by a constant 36 Hz over syllable duration. Thus, for the Mandarin stimuli, symmetric integrality is more difficult to test.

To confirm that the English listeners' perception of the Mandarin stimuli may reasonably be interpreted as integral despite the difference between dimensions in discriminability, these data were subject to a further analysis. In each judgment condition, the amount of integrality English listeners showed was expressed as the ratio of orthogonal to control RTs. A *t* test on these ratios showed that the proportion increase in RT in the orthogonal condition was about the same in the segmental and the suprasegmental judgment conditions ($t = -.338, p > .35$). The proportionately equal increase in RT suggests that the dimensions of the Mandarin stimuli are indeed perceived integrally by the English listeners.

To further examine possible effects of linguistic experience on integrality, the effect of type of suprasegmental information on the degree of dimensional integrality was examined for the Chinese listeners. Since the constant-pitch and Mandarin tones have different lexical functions for the Chinese listeners, it is possible that type of suprasegmental information affects the degree of integrality between the segmental and suprasegmental dimensions for these listeners. Repp and Lin (1990) tested this possibility and found that their Chinese listeners showed greater integrality for the Mandarin-tone stimuli than for the non-Mandarin tones in one of four tasks. To test this in the present experiment, the mean difference in the Chinese listeners' RTs for the control and orthogonal conditions for each test session was calculated. Difference scores reflect the amount of integrality between dimensions. These difference scores were averaged across subjects and judgment conditions for the constant-pitch session and again for the Mandarin session. These mean RTs for the constant-pitch and Mandarin sessions were then compared in a *t* test. The *t* test indicated no significant difference in the amount of integrality that Chinese listeners showed as a function of stimulus set [$t(7) = -.27, p > .39$].

DISCUSSION

Does the perceptual integrality between segmental and suprasegmental information depend on the linguistic function of the suprasegmental information, or does it depend only on the acoustic properties of the two dimensions? Mandarin listeners show mutual and symmetric integrality between suprasegmental and segmental information, even for the constant-pitch stimuli, which are not actual Mandarin tones. Since suprasegmentals are lexically im-

portant in Mandarin Chinese, this integrality shown by the native Chinese listeners is not surprising. Furthermore, since suprasegmentals are not lexically important in English, and in light of Wood's (1974, 1975) results, the finding of asymmetric integrality for the constant-pitch stimuli for the native English listeners is also as expected.

Two aspects of the present set of results, however, are not entirely consistent with an interpretation based on language-specific processing strategies. First, the Chinese listeners showed integrality in their perception of segmentals and non-Mandarin (constant-pitch) suprasegmentals, despite the nonlexical nature of these pitches. Second, the English listeners showed mutual integrality between dimensions in their perception of the Mandarin stimuli, even though Mandarin tones are not lexically relevant in English.

The mutual and symmetric integrality that the Chinese listeners show for the constant-pitch stimuli may have a linguistic basis. Because the suprasegmentals of the constant-pitch stimuli—level pitches—resemble Tone 1 (a high-level pitch) in Mandarin Chinese, listeners might have interpreted the constant-pitch stimuli as Mandarin. The performance of the native Mandarin speakers who judged the stimuli supports this suggestion. One judge interpreted both of the constant pitches as Tone 1 on every trial, and the other interpreted both pitches as Tone 1 on half of the trials. Thus, in the speeded classification task as well, the listeners may have been interpreting the constant-pitch stimuli as Mandarin words. As a further possibility, perhaps the lexical function of suprasegmentals in Chinese makes native listeners process all suprasegmentals, regardless of their degree of resemblance to actual Chinese tones, as integral with their segments. Consistent with this interpretation, in Repp and Lin's (1990) study, Chinese listeners also perceived the non-Mandarin tones (a low rising-falling contour and a low-level tone) integrally with segments.

An explanation based on the lexical function of tone can account for the integral perception shown by the Chinese listeners, but it does not explain the pattern of results for English listeners. Suprasegmental information does not specify lexical items in English as it does in Chinese, yet English listeners perceived the segmentals and Mandarin suprasegmentals in a mutually integral fashion. Why do English listeners show different patterns of processing for Mandarin stimuli and for constant-pitch stimuli? This difference cannot be explained simply on the basis of the acoustic properties of the stimuli, without regard to the linguistic knowledge of the listener, since the Mandarin listeners heard the same sets of stimuli and showed a different pattern of results. Rather, as with the Chinese listeners, perhaps the stimuli that show symmetric integrality do so because of the linguistic informativeness of their suprasegmental structure.

Suprasegmentals are not lexically relevant in English in the way they are in Chinese, yet they convey other kinds of linguistic and paralinguistic information. For example, at the sentence level, changes in pitch signal the relative

prominence of words in the sentence, thus modifying the intended meaning. (For example, "The *dog* has fleas" implies that the dog, not the cat or another animal, has fleas. In comparison, "The dog has *fleas*" suggests that the dog is plagued by fleas, rather than ticks or other pests.) Changes in pitch may also turn statements into questions or convey the doubt or certainty with which a statement is made (see Ladefoged, 1982, chap. 5; see also Bolinger, 1989). The prosody of English conveys affective information (e.g., Cosmides, 1983; Fernald, 1984; Fernald & Kuhl, 1987; Werker & McLeod, 1989). In addition, prosody aids the segmentation and recognition of fluent speech. Listeners who heard sentences spoken in natural or misleading prosody were better able to identify the noun phrases when the prosody was natural (Read & Schreiber, 1982). Even infants show sensitivity to the prosodic cues that mark linguistic boundaries in fluent speech (Jusczyk, 1989). Furthermore, listeners identify words in sentences with normal intonation better than in monotonic sentences (Slowiczek & Nusbaum, 1985).

The various communicative functions that prosody serves may compel native English listeners to attend to fundamental frequency variation in the suprasegmental dimension. For English, an important feature of suprasegmental information may be its dynamic quality. It is the *changes* in intonation that convey the relative prominence of words in an utterance, the affective qualities of speech to infants, and information for the segmentation and recognition of speech. This is consistent with the observation that pitch rises and falls *continuously* throughout an utterance. Constant pitches do not normally occur (see Ladefoged, 1982, chap. 5). Thus, a difference between constant and dynamic pitches in informativeness and naturalness may account for differences in the native English listeners' processing of these types of suprasegmental dimensions. They may process segmental and suprasegmental information integrally only when they expect that both dimensions will provide relevant information for recognition.

In conclusion, the evidence suggests that both Chinese and English listeners must attend to suprasegmental information because they have learned through linguistic experience that this information is important in understanding spoken language. That is, the way listeners perceive the dimensions of the speech signal does not depend simply on the acoustic characteristics of these dimensions. Rather, perception depends on how the structure of the signal interacts with the language-specific processing strategies of the listener. The Chinese listeners in the present study processed all segmental and suprasegmental dimensions on which they were tested in an integral fashion. Since tone is lexically relevant in their native language, perhaps native Chinese listeners have learned to always consider simultaneously information from both segmental and suprasegmental sources in word recognition. For the English listeners, attention to the suprasegmental dimension may benefit language comprehension in more general ways, but only for dynamic pitch con-

tours. Accordingly, these listeners showed an asymmetry in processing the dimensions of constant pitch and segments, but showed mutual integrality in their processing of the Mandarin suprasegmentals and segments. Theories of speech perception and word recognition generally consider only the role of phonetic information in the recognition process (e.g., Klatt, 1980b; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; but see Grosjean & Gee, 1987). However, the present results demonstrate that a complete theory must consider how listeners integrate information from both the segmental and the suprasegmental dimensions of the speech signal in understanding spoken language.

REFERENCES

- BOLINGER, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford: Stanford University Press.
- 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.
- CHOMSKY, N., & HALLE, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- COSMIDES, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 864-881.
- DELATTRE, P. C., LIBERMAN, A. M., & COOPER, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, **27**, 769-773.
- EIMAS, P. D., TARTTER, V. C., MILLER, J. L., & KEUTHEN, N. J. (1978). Asymmetric dependencies in processing phonetic features. *Perception & Psychophysics*, **23**, 12-20.
- FERNALD, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, & R. Golinkoff (Eds.), *The origins and growth of communication* (pp. 5-29). Norwood, NJ: Ablex.
- FERNALD, A., & KUHL, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development*, **10**, 279-293.
- FRY, D. B., ABRAMSON, A. S., EIMAS, P. D., & LIBERMAN, A. M. (1962). The identification and discrimination of synthetic vowels. *Language & Speech*, **5**, 171-189.
- 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.
- GROSJEAN, F., & GEE, J. P. (1987). Prosodic structure and spoken word recognition. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 134-155). Cambridge, MA: MIT Press.
- JUSCZYK, P. W. (1989, April). *Perception of cues to clausal units in native and non-native languages*. Paper presented at the biennial meeting of the Society for Research in Child Development, Kansas City, MO.
- KLATT, D. H. (1980a). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.
- KLATT, D. H. (1980b). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243-288). Hillsdale, NJ: Erlbaum.
- LADEFOGED, P. (1982). *A course in phonetics* (2nd ed.). San Diego, CA: Harcourt Brace Jovanovich.
- MARSLÉN-WILSON, W. D. (1987). Functional parallelism in spoken word-recognition. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 71-102). Cambridge, MA: MIT Press.
- MARSLÉN-WILSON, W. D., & WELSH, A. (1978). Processing interactions during word-recognition in continuous speech. *Cognitive Psychology*, **10**, 29-63.
- MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.

- MILLER, J. L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception & Psychophysics*, **24**, 175-180.
- NEWELL, A. (1975). A tutorial on speech understanding systems. In D. R. Reddy (Ed.), *Speech recognition: Invited papers presented at the 1974 IEEE Symposium* (pp. 3-54). New York: Academic Press.
- READ, C., & SCHREIBER, P. (1982). Why short subjects are harder to find than long ones. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 78-101). Cambridge: Cambridge University Press.
- REPP, B. H., & LIN, H.-B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics*, **18**, 481-495.
- SLOWIACZEK, L. M., & NUSBAUM, H. C. (1985). Effects of speech rate and pitch contour on the perception of synthetic speech. *Human Factors*, **27**, 701-712.
- TOMIAK, G. R., MULLENIX, 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.
- WERKER, J. F., & MCLEOD, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, **43**, 230-246.
- 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 & 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.

NOTES

1. A *correlated* condition is sometimes included in the speeded classification task (e.g., Wood, 1974), in which subjects are presented with repetitions of two stimuli that differ in value for both the target and the nontarget dimensions. In the correlated condition, subjects classify according to a specified target dimension, but since the values on the target and nontarget dimensions are correlated, they may also classify according to variation in the nontarget dimension. Although faster target decisions in the correlated condition may be interpreted as further support for integrality of dimensions, it is possible to get faster recognition in this condition with separable dimensions due to simple redundancy gains. Because results in the correlated condition are difficult to interpret, this condition was not included in the present study.

2. An analysis of variance (ANOVA) indicated that the native English listeners showed a significant main effect of condition. They responded more slowly in the orthogonal condition, when context varied (562 msec), than in the control condition, when context was held constant (506 msec) [$F(1,8) = 20.02, p < .01$]. The main effects of stimulus set (Mandarin vs. constant pitch) and judgment (segmental vs. suprasegmental) were not significant [$F(1,8) = .18, n.s.$, for stimulus set; $F(1,8) = 2.44, n.s.$, for judgment], nor were any of the interactions.

3. Although two dimensions may show integrality, the amount of integrality may be greater in one direction than the other. Asymmetries in integrality may reflect characteristics of the stimulus dimensions or the processing strategies of the listener. To test the symmetry of the integrality effects that English listeners displayed, *t* tests were carried out on the difference scores between subjects' mean orthogonal and control RTs for each judgment condition (segmental, suprasegmental) for the Mandarin stimulus set. The difference scores—the mean increase in RT due to orthogonal variation for each task—reflect the degree of interference in processing from the irrelevant dimension and thus the degree of integrality between dimensions. As expected, since the planned comparisons already demonstrate an asymmetry for the constant-pitch stimuli, English listeners showed a significant difference in the orthogonal effect for segmental versus suprasegmental judgments [$t(8) = 2.9, p < .01$]. For the Mandarin stimuli, a *t* test showed no significant difference in the orthogonal effect between the segmental and suprasegmental judgment conditions [$t(8) = -.56, p > .29$]. Thus, for English listeners, the processing interactions between segmental and suprasegmental dimensions in the Mandarin stimuli were both mutual and symmetric.

4. An ANOVA on the data from Chinese listeners showed that, like the English listeners, the Chinese listeners also responded more slowly in the orthogonal condition, when context varied (759 msec), than in the control condition, when context was constant (659 msec) [$F(1,7) = 25.24, p < .01$]. The main effect of stimulus set approached significance [$F(1,7) = 4.76, p < .07$], indicating that the Chinese subjects were somewhat slower in responding to the Mandarin stimuli than to the constant-pitch stimuli. The main effect of judgment was not significant [$F(1,7) = .25, n.s.$], nor were any of the interactions.

5. The *t* tests showed symmetric integrality for the Chinese listeners. There were no significant differences in the amount of orthogonal interference for either stimulus set [$t(7) = -.55, p > .30$, for constant pitch; $t(7) = -.19, p > .42$, for Mandarin]. Thus, the Chinese subjects showed mutual and symmetric integrality between the segmental and suprasegmental judgment conditions for both the constant-pitch stimuli and the Mandarin stimuli.

(Manuscript received October 18, 1991;
revision accepted for publication July 29, 1992.)