



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Journal of Phonetics 32 (2004) 395–421

Journal of  
Phonetics

[www.elsevier.com/locate/phonetics](http://www.elsevier.com/locate/phonetics)

## Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners

Pierre A. Hallé<sup>a,\*</sup>, Yueh-Chin Chang<sup>b</sup>, Catherine T. Best<sup>c</sup>

<sup>a</sup> *Laboratoire de Psychologie Expérimentale, CNRS-Paris V, France*

<sup>b</sup> *Laboratory of Linguistics, National Tsing Hua University, Taiwan, ROC*

<sup>c</sup> *Wesleyan University and Haskins Laboratories, USA*

Received 28 March 2002; received in revised form 3 March 2003; accepted 5 March 2003

### Abstract

Previous work has not yielded clear conclusions about the categorical nature of perception of tone contrasts by native listeners of tone languages. We reopen this issue in a cross-linguistic study comparing Taiwan Mandarin and French listeners. We tested these listeners on three tone continua derived from natural Mandarin utterances within carrier sentences, created via a state-of-the-art pitch-scaling technique in which within-continuum interpolation was applied to both  $f_0$  and intensity contours. Classic assessments of categorization and discrimination of each tone continuum were conducted with both groups of listeners. In Experiment 1, Taiwanese listeners identified the tone of target syllables within carrier sentence context and discriminated tones of single syllables. In Experiment 2, both French and Taiwanese listeners completed an AXB identification task on single syllables. Finally, French listeners were run on an AXB discrimination task in Experiment 3. Results indicated that Taiwanese listeners' perception of tones is quasi-categorical whereas French listeners' is psychophysically based. French listeners nevertheless show substantial sensitivity to tone contour differences, though to a lesser extent than Taiwanese listeners. Thus, the findings suggest that despite the lack of lexical tone contrasts in the French language, French listeners are not absolutely "deaf" to tonal variations. They simply fail to perceive tones along the lines of a well-defined and finite set of linguistic categories.

© 2003 Elsevier Ltd. All rights reserved.

*Keywords:* Mandarin Chinese; Tones; Categorical perception; Cross-language study

\*Corresponding author. Present address: Université René Descartes, Centre Henri Piéron, 71 Avenue Edouard Vaillant, 92774 Boulogne-Billancourt, France. Tel.: +33-1-55-20-59-34; fax: +33-1-44-52-99-68.

*E-mail address:* [halle@psycho.univ-paris5.fr](mailto:halle@psycho.univ-paris5.fr) (P.A. Hallé).

## 1. Introduction

The literature on the linguistic status of tones in Chinese languages such as Mandarin, as well as many other far-eastern tone languages, is uncontroversial. Tones in these languages may be viewed as phonemic distinctions attached to the syllable at a suprasegmental level. Their main physical correlates are tone-specific fundamental frequency ( $f_0$ ), amplitude (intensity) contours, and—to a lesser extent—duration (cf. Kong, 1987, for Cantonese). Because they serve a lexically contrastive role, they must be perceived by speakers of, say, Mandarin as linguistic elements (perhaps in the same way as vowels and consonants), provided that they are carried by spoken syllables (Van Lancker & Fromkin, 1973; Xu, 1994).

That tone information is used linguistically is supported by a variety of findings. First, there is evidence for a left hemisphere advantage in tone perception by speakers of tone languages. For instance, Van Lancker and Fromkin (1973), using a dichotic task with native speakers of Thai, found a right ear advantage for tones in words but not for “hummed” tones without segmental information (also see Wang, Jongman, & Sereno, 2001, for Mandarin). Gandour and Dardarananda (1983; Gandour, Petty, & Dardarananda, 1988) found that production and identification of Thai words minimally distinguished by tone were significantly impaired in left brain-damaged Thai speakers with aphasia in comparison to both normal subjects and right brain-damaged nonaphasics (also see Yang, 1991). Second, tonal information seems to interact with segmental information in lexical access (Fox & Unkefer, 1985; Lee, 2000). Findings of cross-linguistic differences between native speakers of tonal and nontonal languages have been somewhat less consistent, but generally point to, respectively, linguistic vs. nonlinguistic perception of tones (Fox & Unkefer, 1985; Gandour & Harshman, 1978; Wang, 1976; Lee & Nusbaum, 1993; Lee, Vakock, & Wurm, 1996). Recent neuroimaging studies support this view: Left hemisphere structures are recruited to process tone contours for speakers of Thai or Chinese, whereas right hemisphere structures are engaged for speakers of nontonal languages (Gandour, Wong, & Hutchins, 1998; Gandour, Wong, Hsieh, Weinzapfel, Van Lancker, & Hutchins, 2000; Klein, Zatorre, Milner, & Zhao, 2001). To summarize, tonal information is perceived by native speakers of tonal languages in a linguistically contrastive way, presumably just like segments are.

Our intuition is that Mandarin Chinese tones should be readily categorized by listeners of Mandarin at a “perceptual” rather than a “postperceptual” stage of processing (to use Kolinsky’s, 1998, terminology) in order for word recognition to be achieved efficiently. That is, tone identification is not likely a by-product of word recognition and lexical access: Tones *must* be identified for words to be recognized. Thus, tones should be identified at a prelexical rather than at a (post)lexical level. This might not be the case for speakers of languages in which tones are not used linguistically to distinguish words. In some sense, speakers of English or other stress languages could implicitly exploit stress differences—which include  $f_0$  contour differences, the main physical cue to tone identity—to distinguish words that are otherwise identical. For example, *FORbear* and *forBEAR* (capitalized letters indicate stressed syllables) contain the same segments but make a minimal pair with respect to stress placement. Note that these pairs are rare (a dozen according to Cutler, 1986) and differ from pairs such as *CONvert–conVERT*, which are contrasted for stress placement *and* vowel quality. It is not clear whether speakers of English are sensitive to stress differences at a prelexical level. Indeed, Cutler (1986) found that “*forbear* is a homophone” and that stress “does not constrain lexical access.” Yet, for languages other than

English in which accentual patterns are not reflected at the segmental level, stress may constrain lexical access (Dutch: Cutler & Donselaar, 2001; Japanese: Cutler & Otake, 1999). On the other hand, even in the case of English compared to Mandarin Chinese, Repp and Lin (1990) and Lee and Nusbaum (1993), using the speeded classification paradigm introduced by Garner (1970), found that native speakers of English are sensitive to suprasegmental information in a similar way to Mandarin speakers. This especially holds true for “dynamic”  $f_0$  contours (see Abramson, 1978, for the distinction between “static” and “dynamic” contours). Both English and Mandarin speakers seem to perceive segmental and suprasegmental information *integrally* rather than separately. More strikingly, a recent study by Soto-Faraco, Sebastian-Gallés, and Cutler (2001) shows that speakers of Castillian Spanish are more sensitive than speakers of English to accent mismatch, and presumably at a prelexical level. Thus, different languages induce different sensitivities to stress variations, including  $f_0$  variations.

French provides a likely candidate for a low sensitivity to syllable-level prosodic variations, especially those which involve  $f_0$  and/or intensity contours. French does not employ lexical accent marked by stress placement, and is thus said to be a nonstress language. Indeed, French listeners have been found to suffer from “stress deafness” (Dupoux, Pallier, Sebastian-Gallés, & Mehler, 1997). On the other hand, final lengthening might be considered to be an accentuation mark in French. Final syllable lengthening, which is usually observed in words at the end of prosodic groups, is not otherwise marked in French by  $f_0$  or intensity increases, but rather by a tendency to decrease in final syllable position (Vaissière, 1991). Whether this final lengthening should be considered as a (fixed) accent is a matter of controversy among linguists. In any case, it is not a clear and systematic marker of stress and serves no lexically contrastive function. In effect, in the modern usage of spoken French, both final lengthening and word onset accentuation may occur. As is usually argued, they have a demarcation rather than an accentuation function (Rossi, 1980; Vaissière, 1991). But no language is spoken in a monotone, and French is no exception. French uses *sentence*-level intonation to mark various illocutory modes as well as various moods (see Di Cristo, 1998, for the prosodic correlates of some identified intonation patterns in French; also see Carton, 1974, pp. 91–98). Yet, the use of suprasegmental linguistic distinctions—other than sentential intonation or prosodic group demarcation—by speakers of French appears to be remarkably limited compared to other languages.

### 1.1. The issue of prelexical tone identification

Recent cross-linguistic investigations on tone perception by Cantonese or Mandarin speakers vs. speakers of nontonal languages have focused on the relative role of consonant, vowel, and tone identification in the course of lexical access (Cutler & Chen, 1997; Ye & Connine, 1999; Lee, 2000) with somewhat conflicting conclusions. Tone information logically becomes available later in time than vowel and consonant information since the domain of tone extends at least over the entire syllable rime (Hallé, 1994; Howie, 1974), and this is consistent with the findings of tone vs. segment monitoring studies. Tone identity in Mandarin, however, can be retrieved with reasonable success from rather brief fragments of the syllable’s beginning (Lee, 2000; Whalen & Xu, 1992), with more confusions between tone 2 (“mid-rising”) and tone 3 (“low-dipping”). In running speech, tone contours change in a predictable way according to tone context (Xu, 1994), thus making it possible for tones to be identified *on-line* via anticipation, rather than inferred

regressively from their following context (see Gow, 2001, for a similar view in the case of phonemic assimilation). In short, given the findings gathered thus far and speculations from indirect evidence, it still remains unclear (1) whether speakers of, say, Mandarin categorize tones prelexically and presumably on-line, or postlexically, and (2) whether speakers of nontonal languages, especially nonstress languages such as French, perceive tones differently than speakers of Mandarin. It might thus be premature to make strong claims about the role of tones during on-line lexical access, before we gain more solid basic knowledge about how tones are handled at a basic, perceptual level. In the present report, we try to fill that gap and focus on the prelexical categorization of Mandarin tones by native compared to non-native listeners.

### 1.2. The acoustics of Mandarin Chinese tones

Before reporting the present study, it is necessary to briefly outline the acoustic–phonetic characteristics of tones in Taiwan Mandarin. The tones of Mandarin as spoken in mainland China, as well as those of other Chinese languages, have been described by many investigators (e.g., Chao, 1948, 1968; Ho, 1976; Howie, 1974, 1976; Kratochvil, 1968, 1985, 1998). However, certain points need clarification.

There is a wide consensus for the tone-specific  $f_0$  contours found in citation form. Mandarin tones are traditionally numbered as tones 1–4. The domain of tone—the portion of a syllable that bears a tone-specific  $f_0$  contour—seems to be the syllable rime rather than the entire voiced portion of the syllable (Hallé, 1994; Howie, 1974; but see Xu, 1998, and Xu & Wang, 2001, for the proposition that tone is aligned with syllable onset). Tone 1 is high-level, tone 4 is high-falling, tone 2 is mid-rising. There is some controversy about tone 3, often described as low-dipping or low-falling-rising. Chao (1968) qualified the description of tone 3, positing a “free variant” tone 3, about half in duration of the citation form tone 3, and simply low-falling in contour shape, which he called “half third tone.” The truth of the matter seems to be that tone 3 is pronounced with a rising second half mostly in citation form. In running speech in nonprepausal position, tone 3 is most often pronounced as the “half third tone” (Coster & Kratochvil, 1984; also see Gårding, Kratochvil, Svantesson, & Zhang, 1986; Hallé, 1994; Kratochvil, 1987).<sup>1</sup> The profile of tone 3 observed in normal speech is low-falling, below a speaker’s  $f_0$  register rest level (see Hallé, 1994, for evidence from EMG data), and the final rise observed in citation form may simply reflect a mechanical movement back to rest  $f_0$  level, rather than an intentional linguistic marking.

Taiwan Mandarin tones differ only slightly from mainland China Mandarin tones. One difference that has been proposed is that the tone 2 profile (mid-rising) has a rather pronounced dipping initial portion and a mildly rising final portion in Taiwan Mandarin (Fon & Chiang, 1999). That is, tone 2 of Taiwan Mandarin would not be as clearly rising from a mid-low to a high

<sup>1</sup>A radical view is that of Kratochvil, who proposed that isolated citation forms are, as it were, “demonstration tones” that possibly serve as “names for tones comparable to the spelling names of the letters of the alphabet” (Kratochvil, 1987, p. 258). The spelling metaphor may be especially relevant for tone 3, whose low dipping contour observed in citation form could be viewed as a device to *spell out* the third tone category. On the other hand, tone 3 with a final rise may occur prepausally, as in monosyllabic utterances in spontaneous, spoken Mandarin, such as *hao3* (good!) or *gun3* (get lost!) and so on. In the context of the “tone 3 controversy,” we will keep to a rather factual standpoint: The usual pattern for tone 3 in running speech is the so-called “half third tone” with little or no final rise.

pitch as tone 2 in Beijing Mandarin. Yet, this difference needs to be confirmed by more extensive cross-dialectal research.

Tones also vary in duration. For example, tone 3 has often been described as much longer than the other tones (Ho, 1976; Howie, 1976). But this difference holds true only for tones in citation form. In normal speech, the duration differences between tones are not a reliable cue to their identity (Coster & Kratochvil, 1984; Kratochvil, 1985, 1987, 1998).

Whereas  $f_0$  contour is often taken as the main acoustic cue to tone identity, careful investigations have also noted that intensity or amplitude contour is also an important characteristic of tones (Coster & Kratochvil, 1984; Hallé, 1994; Whalen & Xu, 1992). For example, Coster and Kratochvil (1984) submitted a large corpus of spontaneous speech to discriminant analysis in which tones were described by their  $f_0$  contour, amplitude contour, and duration; they found that the analysis achieved 86.5% correct discrimination when based on  $f_0$  contours only, but still 71.1% correct discrimination when based on the amplitude contours only. Hence, amplitude contours are tone-specific to a large degree. This is confirmed by the unexpectedly high success of listeners at identifying tones in synthetic “syllables” in which only amplitude information has been maintained and spectral information has been replaced with white noise (Whalen & Xu, 1992).

### 1.3. Categorical perception of tones

Categorical perception may offer a particularly useful and sensitive approach to address the issue of how Mandarin Chinese tones are perceived and possibly categorized by native speakers as compared to speakers of a language such as French, which is utterly nontonal at the contrastive phonological level. Both Mandarin and nontonal languages use prosodic variations for various purposes. But Mandarin uses tonal variations in a phonologically contrastive way at the lexical level, whereas French, for example, instead uses prosodic variations at the sentence level in a rather loose way to cover an *open* set of communicative intentions. In the present study, we compared the performance of Mandarin and French listeners on the perception of Mandarin tones in a classical categorical perception investigation that employed discrimination and categorization of synthetic speech continua. As we have argued elsewhere (Hallé, Best, & Levitt, 1999), the use of carefully controlled continua—as opposed to natural tokens—allows for the detection of fine-grained differences in performance among listener groups. In other words, both the materials and the subject populations we use here tend to maximize the cross-linguistic differences we expect to find: Mandarin listeners should process tones as contrastive linguistic categories, displaying some degree of categorical perception, whereas French listeners are not expected to do so.

There exist a few pioneering studies of the categorical perception of tones. Abramson (1979) claimed that tone perception in Thai is not categorical. But he used a strict criterion to determine categoricity, at a time when the categorical perception found for stop consonants typically involved both steep slopes in categorization functions and marked peaks at the category boundary in discrimination functions. These two facets of speech continua perception were considered to be the “signature,” as it were, of categorical perception (but see Massaro, 1987; Massaro & Cohen, 1983, for a critical view of the very notion of categorical perception). Abramson (1979) used a continuum of 16 level tones (constant  $f_0$  contours) between 92 and 152 Hz imposed on the syllable

[kha:]. This continuum encompassed three of the five Thai tones: the three static low, mid, and high tones. Most of the Thai participants did a reasonably good job at parsing the continuum into these three tone categories. Their discrimination performance, however, was high throughout the continuum and did not show clear peaks at the presumed boundary locations. Note that static tones might be expected to yield lower categoricity, just like steady-state vowels as compared to vowels between consonants in a more natural dynamic context (Stevens, 1966). The perception of dynamic tones is, perhaps, intrinsically more categorical. Compatible with that possibility, an 11 step continuum with the syllable [i], varying between a 135 Hz level tone and tones rising in a linear fashion up to 135 Hz from a starting  $f_0$  of 105–132 Hz, revealed a typical pattern of categoricity for Mandarin listeners but not for English listeners (Chan, Chuang, & Wang, 1975, reported in Wang, 1976). This study, however, involved a very limited number of participants. Based on these findings, Wang (1976) claimed that tone perception in Mandarin Chinese was clearly categorical. A different kind of evidence also provides support for categorical perception of tones by Mandarin listeners but not by English listeners: Stager and Downs (1993) reported that Mandarin listeners are *less* sensitive than English listeners to small  $f_0$  contour variations in a same-different discrimination task. This seems rather counterintuitive at first sight but is consistent with the notion that Mandarin listeners must ignore irrelevant tonal variations (i.e., within-category variations) in order to efficiently categorize  $f_0$  contours into tones. English listeners do not have to do so.

No further studies using tone continua—to our knowledge—have reconsidered the issue of tone categorization until our recent endeavor with Taiwan Mandarin listeners (Chang & Hallé, 2000, published in Chinese). The results of that study suggest a gradient in categoricity, with tones roughly similar to vowels in degree of categoricity in perception. A possible reason for such similarity is that both tones and vowels extend over a similar time range and thus contain extensive information which, however, requires time to become fully available to perception. This interpretation is consistent with recent observations on the role of tone and vowel in lexical access (Cutler & Chen, 1997; Ye & Connine, 1999), pointing to the late availability of tone and vowel information. Another reason is the overlap between contrasting tone categories or vowel categories, due partly to contextual variation and partly to the relational (as opposed to absolute) nature of the acoustic information that defines tone and vowel categories (for tones, see Connell, Hogan, & Rozsypal, 1983; Fox & Qi, 1990; Leather, 1983; Xu, 1994).

In this article, we first present our previous findings with Taiwanese Mandarin listeners. We then report categorization experiments conducted with both Taiwanese and French subjects, and further discrimination experiments with French subjects. These experiments addressed three questions: (1) How categorical is the perception of tones by native speakers of Mandarin Chinese? (2) To what extent will (naïve) French listeners be sensitive to tone contour differences? (3) Is there some indication that French listeners perceive such differences in a linguistic way rather than in a merely psychophysical fashion?

## 2. Experiment 1: Mandarin Chinese listeners' identification and discrimination of tones

In this section, we report the results obtained with Taiwan Chinese participants on identification and discrimination tasks using three tone-to-tone continua (Chang & Hallé,

2000). In this part our aim is to provide a clear picture of how categorical the perception of Mandarin tones by native listeners is, looking for increases in discrimination performance and in identification accuracy at putative category boundaries.

## 2.1. Method

### 2.1.1. Participants

Fifteen participants from the National Tsing Hua University in Taiwan (aged 22–30 years) participated for a small amount of money in three experimental sessions: one session per continuum, each involving an identification test followed by a discrimination test.

### 2.1.2. Speech materials

The materials used in all the experiments reported below were derived from natural utterances of Mandarin Chinese syllables. Tone continua were constructed using naturally produced syllables for the continuum endpoints. The intermediate contours were obtained via interpolation between endpoints instead of using less natural level  $f_0$  contours or linearly rising (or falling) contours. Because intensity contours are tone-specific, though to a lesser extent than  $f_0$  contours, the continua that we created integrate both the intensity and  $f_0$  dimensions. We had a male native speaker of Mandarin Chinese (Taiwan Mandarin) pronounce the target syllables /pa/, /pi/, and /kwo/ at each of the four tones, within the carrier sentence “yi ge X zi” (“one character X”) where “X” stands for a given target syllable. Each sentence token was pronounced at least four times. The use of three (segmentally simple) different syllables varying in rime and/or onset was intended to avoid overly monotonous tasks. No predictions were made as to the possible differences in performance among these three syllables. Importantly, all three syllables could combine with any of the four tones. That is, there was no “nonword” in the materials. For each of the 12 syllable  $\times$  tone combinations, there were several homophones, some of which were high in frequency of occurrence (above 100 per million according to Wang et al., 1986). This was another motivation for the choice of these three syllables. Indeed, given the rather large heterogeneity in frequency of Mandarin Chinese words for a given syllable across the four tones—with a number of distributional gaps—it is not a simple matter to meet the requirements of both segmental simplicity and variation, and of frequency homogeneity. The frequency data, together with the most frequent morpheme for each syllable  $\times$  tone combination, are provided in the Appendix. Twelve sentences, each corresponding to a syllable and tone combination (three syllables  $\times$  four tones), were retained for their overall homogeneity in global intonation, intensity, and articulation rate. For each target syllable, three continua were constructed: tone 1–tone 2, tone 2–tone 4, and tone 3–tone 4. One motivation was to include quasi-static-to-dynamic tone continua in both the high range of  $f_0$  (t1–t2) and the low range of  $f_0$  (t3–t4) in order to get data comparable to that obtained by Wang (1976) who used a continuum between a linearly rising  $f_0$  ramp and a constant  $f_0$  contour. We added the t2–t4 dynamic-to-dynamic continuum whose endpoint contours are dramatically opposed as rising vs. falling.

Each continuum proceeded through eight steps from one endpoint to the other. The original sentences used to create the various endpoint syllables were measured for  $f_0$  and intensity. For each continuum, one of the two endpoints was chosen as the initial endpoint speech signal from which all the eight steps were derived: t1 for the t1–t2 continuum, t2 for t2–t4, and t3 for t3–t4.

For each continuum, a stylized  $f_0$  contour (a smoothed version of the original contour) was imposed on the initial endpoint syllable (e.g., /pa2/ for /pa2-/pa4/), using a pitch-scaling algorithm similar to the time-domain “pitch synchronous overlap add” (PSOLA) method (Moulines & Laroche, 1995), which is known for the high degree of naturalness achieved (cf. Semal, Demany, Ueda, & Hallé, 1996, footnote 3, for details on our implementation). The final endpoint of the continuum was obtained by imposing on the same syllable the stylized and time-adjusted  $f_0$  contour of the other endpoint original syllable (e.g., /pa4/ for /pa2-/pa4/). The resulting waveform was further modified by changing its intensity contour to that of the final endpoint original syllable. For example, the /pa2-/pa4/ continuum started from a modified version of the original /pa2/ syllable (in which the original  $f_0$  contour was smoothed) and ended in another modified version in which both  $f_0$  and intensity contours were replaced by those of the original /pa4/ syllable (time-adjusted and smoothed). The remaining six intermediate stimuli were obtained by interpolating, at each time point, both  $f_0$  on a log frequency scale and intensity on a decibel scale. For each of the eight steps of each continuum, sentence-level  $f_0$  and intensity contours were imposed on the entire carrier sentence including the target syllable, instead of cross-splicing the carrier sentence and modified target syllable. This option was chosen in order to avoid potential differences and discontinuities in speech quality between the steps of the continua, including the endpoint steps. It was important to maintain a homogeneous quality at the sentence level because we used whole sentences in the identification tests designed for Taiwanese participants. For all the other tests, we used only the target syllables excerpted from their sentence context. The eight  $f_0$  contours for each continuum with the syllable /pi/ are shown in Figs. 1A–C. The intensity contours for the original /pi/ syllables in the four tones, shown in Fig. 2, illustrate that intensity contours are strongly tone dependent and tend to correlate with  $f_0$  contours. Durations, though tone-specific to some extent, stay within a narrow range of variation.

The naturalness of the synthesized sentences was first checked informally by the authors. We further asked 14 native speakers of Taiwan Mandarin to “rate their naturalness in terms of speech quality” on a 1–5 scale. Our concern was that large modifications in  $f_0$  and amplitude might alter the naturalness of synthesized speech, especially for the final endpoint sentences. The ratings

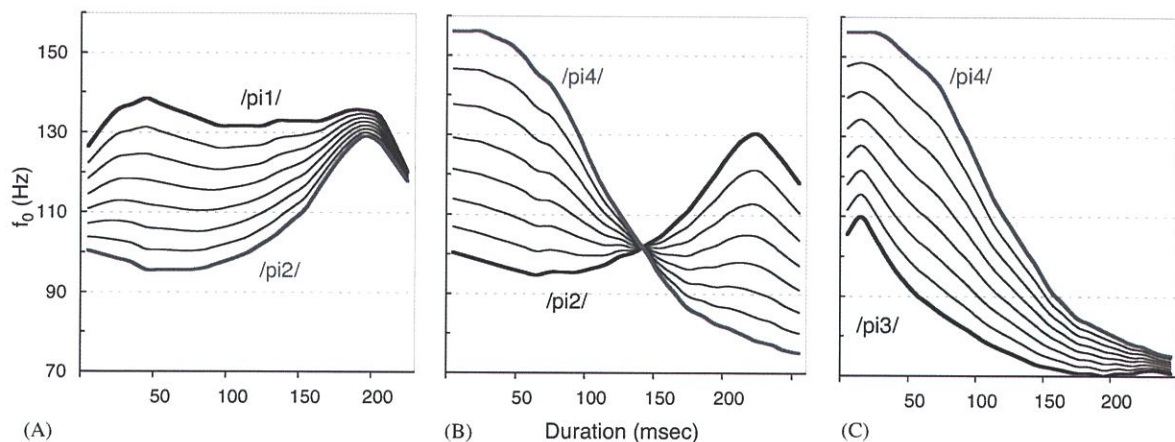


Fig. 1. (A–C). Tone contours for /pi/ in the three continua: (A) in t1–t2, (B) in t2–t4, and (C) in t3–t4.



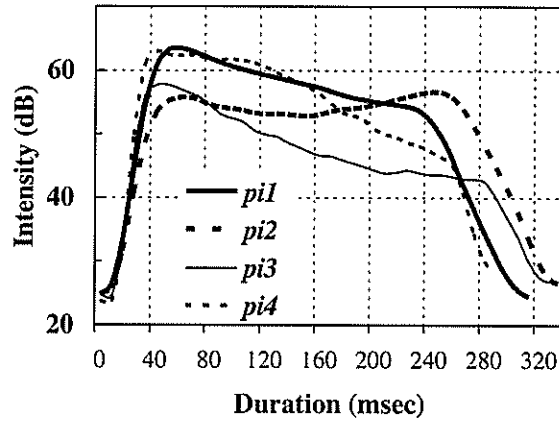


Fig. 2. Intensity contours of the syllable /pi/ in each of the four tones (original stimuli).

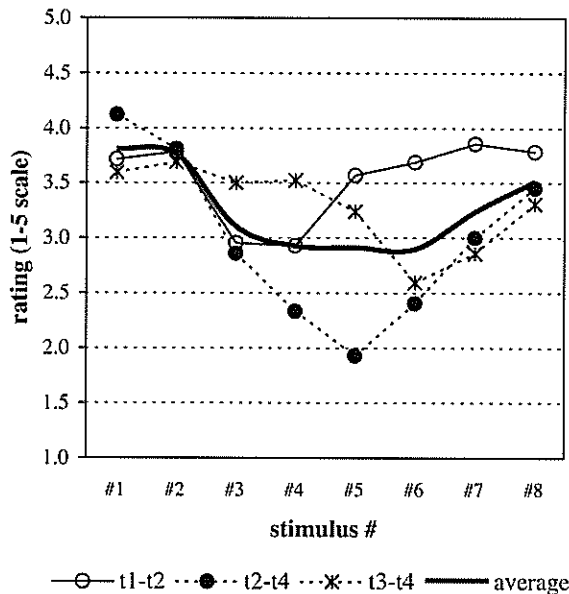


Fig. 3. “Speech quality ratings” of the stimuli (whole sentences), according to stimulus number and to tone-continuum.

(Fig. 3) showed that the final endpoint sentences were not rated lower than the initial endpoint sentences (3.52 vs. 3.81), although the latter are very close to the original sentences and the former had undergone rather drastic changes in both  $f_0$  contour (Fig. 1) and amplitude contour (Fig. 2). The U-shaped curves in Fig. 3 show that *intermediate* rather than final endpoint sentences were rated the lowest. Variations in quality rating thus did not reflect the extent of pitch-scaling and amplitude transformation.<sup>2</sup>

<sup>2</sup>These observations were substantiated by statistical analyses. An ANOVA analysis was run with continuum step and continuum type as between items variables (ratings were pooled across subjects and syllable types). These two

### 2.1.3. Identification tests

For each continuum, participants were presented with 20 repetitions of each sentence stimulus in the test phase. This made a total of 480 trials (three syllables  $\times$  eight steps  $\times$  20 repetitions) presented in quasi-random order, and blocked by 16 trials. The test phase was preceded by a training phase of 32 trials. In this phase, only four steps (the two endpoints and steps #3 and #6) for the syllables /pa/ and /kwo/ were used. The intertrial interval was set to 3.2 s (training) or 2.8 s (test) and the interblock interval to 8 s.

Participants were asked to label the tone of the target syllable with a forced choice between two Chinese characters representative of the two endpoint tones for each continuum (those characters shown in the Appendix). They received written and oral instructions in Mandarin. One participant did not complete the identification test in the t2–t4 session (but was successfully tested on the discrimination test). The data for the corresponding continuum thus involve 14 instead of 15 participants.

### 2.1.4. Discrimination tests

Participants were run on an AXB two-step discrimination test, where the stimuli were the target syllables excised from the carrier sentences used in the identification test. The AXB trials had four possible combinations (AAB, ABB, BAA, and BBA). For each continuum, there were 6 A–B pairs (1–3, 2–4, 3–5, 4–6, 5–7, and 6–8). In the test phase, participants received 360 trials (three syllables  $\times$  six pairs  $\times$  four combinations  $\times$  five repetitions) presented in quasi-random order, and blocked by 20 trials. The test phase was preceded by a training phase of 32 trials, including four pairs (1–3, 3–5, 4–6, and 6–8) in all four AXB combinations for the syllables /pi/ and /kwo/. The interstimulus interval was set to 1 s, the intertrial interval to 3.2 s (training) or 2.8 s (test) and the interblock interval to 8 s.

For each trial, participants had to circle either the number “1” (X=A) or the number “3” (X=B) on a prepared answer sheet. As for the identification test, the instructions were given in Mandarin.

## 2.2. Results and discussion

### 2.2.1. Identification

Identification accuracy and category boundary locations were assessed by means of probit analyses of individual identification curves: They were estimated as the slopes and intercepts, respectively, of the Gaussian distribution functions fitted to the raw data. The particulars of the technique used—short ogive fitting—are described elsewhere (e.g., Best & Strange, 1992; Hallé et al., 1999). Group identification curves are shown in Figs. 4A–C. Table 1 summarizes the slope and intercept data pooled across participants for each continuum and syllable. Analyses of

(footnote continued)

variables both had a significant effect and interacted significantly (all  $ps < 0.01$ ), indicating that (1) ratings significantly varied along continua (the curves in Fig. 3 are indeed all U-shaped), and (2) ratings varied differently for each tone continuum. In particular, the low point in the U-shaped curves varied in both rating value and location: step 3–4, 5, and 6, approximately, for t1–t2, t2–t4, and t3–t4 continua, respectively. Importantly, ratings at step 1 (initial endpoint) were not higher than those at step 8 (final endpoint): 3.81 vs. 3.52,  $F(1, 6) = 2.40$ ,  $p = 0.17$ . The mid-continuum region of the U-shaped curves (steps 3–6) was significantly lower than the two endpoints (steps 1 and 8),  $F(1, 6) = 34.69$ ,  $p = 0.0013$ .

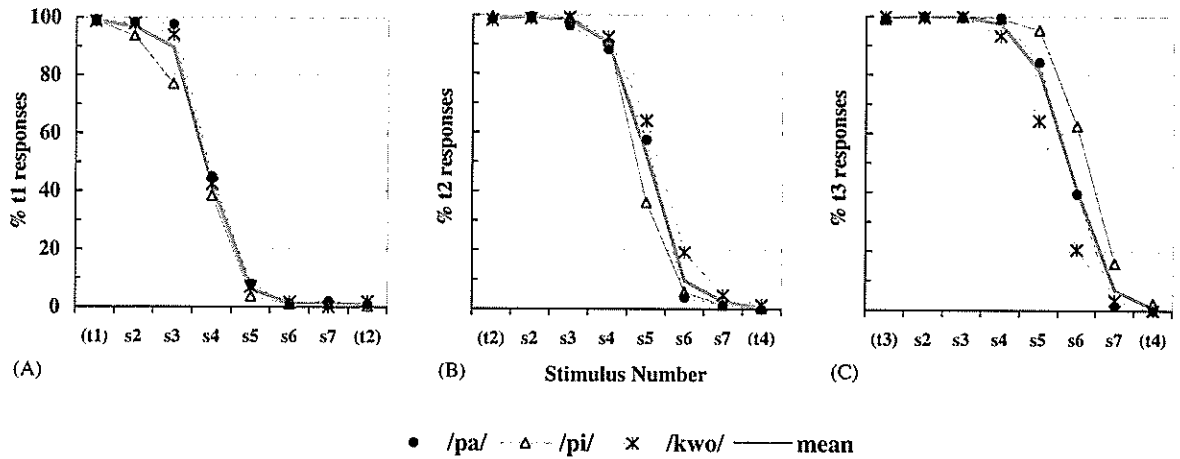


Fig. 4. (A–C). Identification curves for Taiwanese participants in the three continua: (A) in t1–t2, (B) in t2–t4, and (C) in t3–t4 (tone continua in the sentence context “yi ge X zi”).

Table 1  
Intercept and slope for the Taiwanese participants’ identification data (using sentences)

Continuum	Intercept (stimulus number)				Slope (1/SD values)			
	Syllable type				Syllable type			
	/pa/	/pi/	/kwo/	Mean	/pa/	/pi/	/kwo/	Mean
t1–t2	3.91	3.58	3.83	3.77	1.90	2.06	1.83	1.93
t2–t4	4.91	4.76	5.29	4.99	2.28	2.18	1.91	2.12
t3–t4	5.73	6.22	5.30	5.75	2.23	1.60	1.84	1.89
Averages	4.85	4.85	4.81	4.84	2.13	1.95	1.86	1.98

Slopes are actually 1/SD, SD being the standard deviation of the Gaussian functions fitted to the data.

variance on the slopes and on the intercepts were run, with tone continuum and syllable type as within-subject variables. The slopes did not differ significantly across tone continua,  $F(2, 41) < 1$ , but did differ for syllable type: Slope tended to be steeper for the /pa/ syllable type than for /pi/ or /kwo/,  $F(1, 41) = 4.88$ ,  $p = 0.031$ . In contrast, the intercepts differed significantly with tone continuum,  $F(2, 41) = 63.9$ ,  $p < 0.0001$ , but not with syllable type,  $F(2, 82) < 1$ . The boundary was smallest for t1–t2 (closer to tone 1: stimulus number 3.77), larger for t2–t4 (closer to tone 4: stimulus number 4.99) and largest for t3–t4 (closer to tone 4: stimulus number 5.75). These differences among continua were significant at least at the  $p < 0.001$  level ( $t$ -tests).

These variations did not systematically reflect differences in morpheme frequency. While the boundary for /kwo3/-/kwo4/ corresponded to more tone 4 judgments than for /pi/ and /pa/, in line with the higher frequency of /kwo4/ than /kwo3/, the other differences did not lend themselves to interpretations in terms of differential frequencies. Such effects have been obtained for extreme cases of words vs. nonwords (Fox & Unkefer, 1985). Therefore, as we expected, the

frequency differences in our materials were probably not sufficient to induce lexical frequency biases. The identification curves had a relatively steep slope at crossover. The steepness, however, should be estimated with respect to the acoustic differences from one step to another. Given the way the materials were constructed, this difference was constant in terms of  $\log f_0$  mean absolute difference between two successive contours in the continuum. There were, however, differences between continua. To give a rough idea of the step-to-step differences at boundary crossover, a 100% change in category judgment for the syllable /pi/, for example, corresponded to a mean absolute difference of 8 Hz for t1–t2, 14 Hz for t2–t4, and 12 Hz for t3–t4 (RMS differences 9, 15, and 14 Hz, respectively). These values are well above the frequency difference limens of about 1–2 Hz reported by Stagray and Downs (1993), who used level tones centered on 125 Hz, a frequency which approximates a spoken fundamental.

### 2.2.2. Discrimination

The results pooled across participants are plotted in Figs. 5A–C. As can be seen, the curves are bell-shaped: The performances are high, averaging 88% correct discrimination, with no significant differences in performance between continua or between syllable types; no sharp peak emerges as in the case of discrimination within stop consonant continua, only shallow, broad peaks. Nonetheless, analyses of variance conducted on these data show that pairs 3–5 and 4–6 yield significantly more correct discrimination than the adjacent pairs (e.g., pairs 2–4 and 5–7). For the t1–t2 continuum, correct discrimination is highest at pair 3–5 where it reaches 94%, which is significantly higher than 90% at pair 2–4, or 89.9% at pair 4–6 ( $F(1, 42) = 10.93$ ,  $p = 0.002$  and  $F(1, 42) = 10.59$ ,  $p = 0.0023$ , respectively). For the t2–t4 continuum, a fuzzy maximum is reached at the pairs 3–5 and 4–6 (89% and 89.9%, respectively), significantly higher than the discrimination level reached in the adjacent pairs (86.6% at pair 2–4 and 86.4% at pair 5–7:  $F(1, 42) = 5.04$ ,  $p = 0.029$ , and  $F(1, 42) = 8.96$ ,  $p = 0.0046$ , respectively). A similar outcome obtains for the t3–t4 continuum, that is, a fuzzy maximum at pairs 3–5 and 4–6 (89.8% and

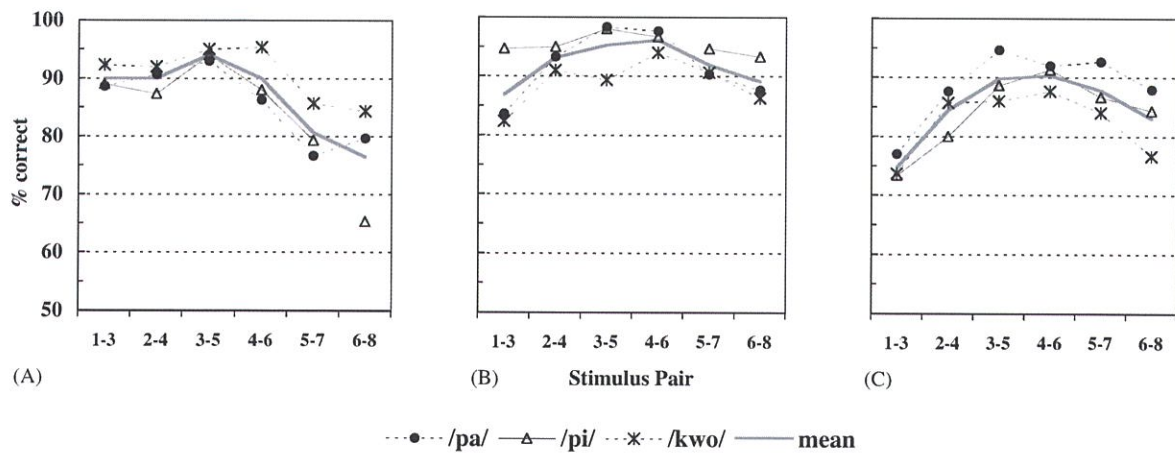


Fig. 5. (A–C). Two-step discrimination curves for Taiwanese subjects in the three continua: (A) in t1–t2, (B) in t2–t4, and (C) in t3–t4.

90.3%, respectively), higher than the level reached in the adjacent pairs (84.4% at pair 2–4 and 87.8% at pair 5–7:  $F(1, 42) = 13.32$ ,  $p = 0.0008$ , and  $F(1, 42) = 3.30$ ,  $p = 0.073$  [marginal], respectively). For the two latter continua, the “peak” of discrimination tends to be more on the side of pair 4–6 than of pair 3–5. Consequently, there is a trend for the peak locations to be consistent with the boundary locations found in the identification tests: The boundary location for t1–t2 is close to stimulus 4, whereas it instead falls between stimuli 5 and 6 for t2–t4 and t3–t4. Interestingly, these locations also tend to coincide with the troughs of the U-shaped rating curves (Fig. 3): around stimuli 3–4 for t1–t2, 5 for t2–t4, and 6 for t3–t4. The lower ratings at *mid-continuum* thus presumably reflect target syllable ambiguity, not natural speech quality per se.

The conclusions drawn from these data are that tone perception by Chinese listeners is categorical (steep slope at category boundary in identification curves) but yet is not as categorical as in the case of stop consonants (shallow, though significant, peaks in the discrimination curves). As we suggested in the Introduction, assuming a gradient of categoricity for various types of speech segments, the categorization of tones is similar to that of vowels. At any rate, the pattern of data we obtained are thus far in line with the notion that tones are processed by Mandarin listeners in a similar way as other segmental contrasts, at least in these basic tests of identification and discrimination along continua.

A more stringent test of how categorical the perception of tones by Mandarin listeners is requires a comparison with their perception by listeners of a nontonal language. As we argued earlier, French presumably provides a suitable nontonal reference language because it does not use lexical prosody contrasts. We thus turned to the question of whether French listeners would exhibit a different pattern of discrimination and categorization performance than listeners of Mandarin Chinese. It could be the case that French listeners perceptually separate tonal information from segmental information, and therefore might perceive tonal variations with reasonably good accuracy. However, it might be that suprasegmental and segmental information intrinsically *cannot* be separated into different streams. In this case, French listeners should hear /pa2/ and /pa4/ as well as all the intermediate steps in the /pa2/–/pa4/ continuum as prosodic variations on the syllable /pa/, yet be unable to assign prosodically functional labels drawn from a closed set of prosodic labels. They might therefore show lower sensitivity overall to within-continuum differences. At any rate, their sensitivity should not be driven by the linguistic value of the stimuli. In other words, French listeners should not show signs of linguistic categorization.

To evaluate these possibilities, we ran French listeners (naïve with respect to Mandarin or to tone languages in general) on the materials described earlier, using AXB identification and discrimination tasks. A new set of Taiwan Chinese participants was run on the AXB identification task for comparison with French participants.

### 3. Experiment 2: Mandarin Chinese and French listeners' AXB identification performance

Because French listeners who are not acquainted with Mandarin tones cannot possibly label them, we used an AXB identification procedure to test French participants' tone identification performance (see Best, Morongello, & Robson, 1981). It should be noted that with this procedure, performances might reveal rather subtle deficits in (linguistic) categorization: The identification curves may well exhibit the usual sigmoid shape and encompass the full range

of 0–100% identification for each endpoint but yet show “defective” (e.g., non-native-like) crossover locations and slopes.

### 3.1. Method

#### 3.1.1. Materials and design

The speech materials were the same as those in Experiment 1. However, we did not use the syllable /kwo/ in this experiment in order to keep the experiment duration within reasonable limits. (The results on /kwo/ had not differed from /pa/ or /pi/ for Mandarin Chinese listeners in any case.) Participants were run in three sessions, one per continuum. For each continuum, A and B in the AXB identification test corresponded to the two endpoints in the two possible orders; X varied from one endpoint to the other along the eight steps of the continuum. The test phase comprised a total of 160 trials (eight steps  $\times$  two orders  $\times$  two syllables  $\times$  five repetitions). It was preceded by a training phase of 32 trials with only four steps (the two endpoints and steps 3 and 6) and two repetitions of each AXB triplet. The interstimulus interval was set to 1 s, the intertrial interval to 3.2 s (training) or 2.8 s (test) and the interblock interval to 12 s (including a 0.5 s warning tone at the beginning of each block). Participants were instructed (in their native language) to press one of two response buttons labeled “1” and “3” for each trial, according to whether they thought the second stimulus (X) sounded more like A than B or vice versa. The response times were recorded from the onset of the third stimulus (B). Participants were instructed to answer as quickly as possible, whenever they were confident in their response. This left open the possibility for participants to respond even before they heard the third stimulus of a triplet, using a same/different strategy on the AX pair. Such a strategy, however, presumably reflects sufficient confidence in judging AX pairs. When participants are not very confident in their response, they indeed have to listen to the third stimulus B. Potentially, then, this lack of confidence can be dramatically reflected by much longer response times.

#### 3.1.2. Participants

Fourteen French students at Paris V University, aged 20–31 years, and 14 Chinese students from Taiwan, aged 24–32 years, participated in the experiment. None of the French participants had ever been exposed to Mandarin or any other tone language. All the participants reported normal hearing and speaking. They participated in the experiment either voluntarily, for a small amount of money, or for course credit.

### 3.2. Results and discussion

The data for each subject was fitted to a short ogive Gaussian curve (cf. Best & Strange, 1992; Hallé et al., 1999). This yielded individual data of intercept and slope at crossover for each continuum and each syllable type. Table 2 summarizes these data for French and Taiwanese participants. The identification curves for the three continua with the syllable /pi/, averaged for each language group, are shown in Fig. 6. Analyses of variance were conducted on the intercept and on the slope data, with Language group (Chinese vs. French) as a between-subject variable, Syllable type (/pa/ vs. /pi/) and Continuum (t1–t2, t2–t4, and t3–t4) as within-subject variables. Syllable type had no significant effect in any of the analyses and will not be further discussed.

For intercepts, Language group had a significant effect,  $F(1, 26) = 8.65$ ,  $p = 0.0067$ . The intercept for French participants averaged 4.62 ( $SE = 0.54$ ), farther from the left endpoint than for Chinese participants ( $M = 4.28$ ,  $SE = 0.56$ ). Both groups showed variations in intercept location across the three continua. However, the intercepts for French participants tended to cluster around 4.5, the exact center of an eight-step continuum, whereas those for Taiwanese participants fell to the left of the center, except for the t2–t4 continuum. Indeed, intercept did not

Table 2  
AXB identification tests: Intercept and slope data for the French and Taiwanese participants

Continuum	Intercept (stimulus number)			Slope (1/SD values)		
	Syllable type			Syllable type		
	/pa/	/pi/	Mean	/pa/	/pi/	Mean
<i>French participants (N = 14)</i>						
t1–t2	4.27	4.20	4.23	1.20	1.06	1.13
t2–t4	4.69	4.85	4.77	1.11	1.17	1.14
t3–t4	4.92	4.82	4.87	1.05	0.97	1.01
Averages	4.63	4.62	4.62	1.12	1.07	1.10
<i>Taiwanese participants (N = 14)</i>						
t1–t2	4.00	4.08	4.04	1.71	2.57	2.14
t2–t4	4.62	4.69	4.66	2.05	2.18	2.11
t3–t4	3.91	4.38	4.15	1.66	1.65	1.66
Averages	4.18	4.39	4.28	1.81	2.13	1.97

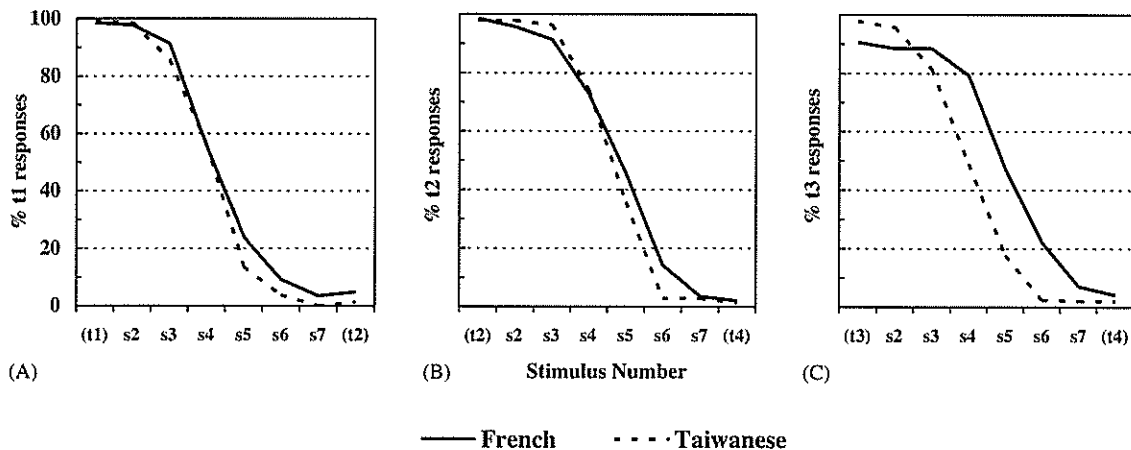


Fig. 6. (A–C). Identification curves of French and Taiwanese participants in the AXB identification task (continua with the syllable /pa/).

differ overall from 4.5 for French,  $t(41) = 1.54$ ,  $p = 0.13$ , whereas it did for Taiwanese,  $t(41) = 2.52$ ,  $p = 0.015$ . This is also shown by a significant Language group  $\times$  Continuum interaction,  $F(2, 52) = 3.85$ ,  $p = 0.027$ .

Slopes at crossover were significantly steeper for Taiwanese than for French participants, for all continua: 1.97 vs. 1.10,  $F(1, 26) = 39.48$ ,  $p < 0.0001$ . Thus, both the intercept and slope data point toward more categorical perception for Taiwanese listeners, and toward more psychophysically based perception for French listeners.

The response time data are illustrated in Fig. 7 for the /pi2/–/pi4/ continuum. As can be seen in this figure, RTs were much longer in the intermediate region, reaching a peak value for X in the AXB categorization trials between the fourth and fifth stimulus in a continuum. However, Fig. 7 suggests that the prominence of this peak was greater for Taiwanese than for French participants. If substantiated by the data, this would support the interpretation that Taiwanese listeners show a more categorical pattern than French listeners. The Taiwanese participants' trouble with responding to the ambiguous stimuli at mid-continuum should be due not only to psychophysical perceptual ambiguity but also to linguistic categorical ambiguity. In order to quantify this group difference, we had to reduce the data in a relevant way. For each identification curve, we retained two RT values: the "peak" value (around mid-continuum) and the average of the two endpoint values, the "edges" value (which generally corresponded to the shortest RTs as can be seen in Fig. 7). Table 3 summarizes the results of this data reduction. Analyses of variance were run on these data, with Language group as a between-subject variable, and Measure ("peak" vs. "edges"), Continuum, and Syllable type as within-subject variables. They showed that Taiwanese and French did not differ significantly with respect to peak RTs,  $F(1, 26) < 1$ , but differed with respect to edges RTs,  $F(1, 26) = 6.87$ ,  $p = 0.014$ . The differential between peak and edges, which is a direct measure of peak prominence, was larger for Taiwanese than for French subjects. This is

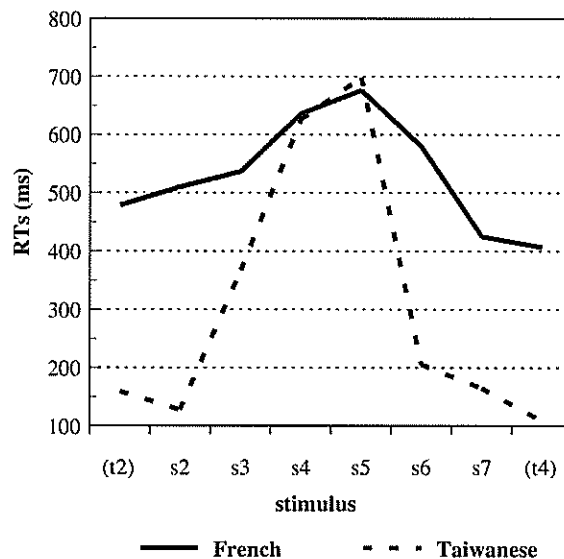


Fig. 7. Response times of French and Taiwanese participants in the AXB identification task (/pa2/–/pa4/ continuum).



Table 3

Saliency of RT peaks at crossover, as estimated by maximum (“Peak”) and endpoint values (“Edges”) in the three tone continua (standard deviations between brackets) for the /pi/ and /pa/syllables

	French Ss		Taiwanese Ss	
	Peak	Edges	Peak	Edges
/pi/	742 (229)	492 (196)	667 (471)	151 (490)
/pa/	754 (207)	487 (187)	768 (480)	193 (499)
Mean	748	490	717	172

shown by the significant interaction between Language group and Measure,  $F(1, 26) = 15.76$ ,  $p = 0.0006$ . The large variability in the Taiwanese participants’ RT data (see Table 3), was mainly due to the fact that some, but not all, of the Taiwanese participants responded even before the third stimulus in AXB triplets, whereas French subjects mainly responded after that stimulus, even for the “easiest” trials where X was *physically* identical to A or to B. To summarize, the peak RT prominence was more marked for Taiwanese participants but the data also tell us this is because Taiwanese were faster by some 320 ms on average for the endpoint trials involving a full category difference, rather than because they were slower at mid-continuum (their peak RTs were not longer than those of French participants, as is seen in Table 3, or Fig. 6).

The results described thus far all indicate that French and Taiwanese listeners’ performance at categorizing tones is qualitatively different. In these AXB identification tests, French listeners can tell whether an intermediate tone is closer to one endpoint than to the other with reasonable success. Yet their accuracy for the critically ambiguous contours is about half that of Taiwanese listeners. Their categorization “boundaries” correspond to the physical centers of the three continua. Finally, they fail to exhibit as sharp a peak RT prominence at the category boundary as Taiwanese listeners do, and this is due to their longer RTs on the endpoint trials; put another way, they have more trouble than Taiwanese at categorizing unambiguous tones. Together, these findings provide converging evidence that the French listeners’ identification performance in the AXB setting reflects psychophysical rather than linguistic sensitivity to similarities and differences in  $f_0$  and intensity contour, whereas the Mandarin listeners’ performance is “boosted” by the phonological value assigned to tone in Mandarin Chinese.

#### 4. Experiment 3: French listeners’ AXB discrimination performance

An even more direct kind of evidence for the nonlinguistic treatment of Mandarin Chinese tone contours by French listeners could be obtained from their discrimination performance along the tone continua. In the discrimination test reported in Experiment 1, Mandarin listeners’ sensitivity to  $f_0$  and intensity contour differences appeared to be biased by the tone category they perceived

in any given contour. This should not be the case for French listeners, for whom there is no phonological significance for any of the contours within a continuum. In order to test this prediction, we had French listeners perform a discrimination task on the three tone continua.

#### 4.1. Method

##### 4.1.1. Design

Participants were run in three sessions: one session per continuum. In each session, they received an AXB two-step discrimination test, identical to that used for Mandarin listeners, except that the materials were limited to the two syllables /pa/ and /pi/ (as in the AXB identification test), and each AXB triplet was repeated only three times (instead of five) in the test phase. (Given the four possible combinations AAB, ABB, BAA, and BBA, this design yielded 12 observations per stimulus pair for each participant.) In the test phase, participants thus received 144 trials (two syllables  $\times$  six pairs  $\times$  four combinations  $\times$  three repetitions) presented in quasi-random order, blocked by 12 trials. The preceding training phase comprised 32 trials, including four pairs (1–3, 3–5, 4–6, and 6–8) in all four AXB combinations for the syllables /pi/ and /pa/. The interstimulus interval was set to 1 s, the intertrial interval to 3.2 s (training) or 2.8 s (test), and the interblock interval to 12 s. Participants received exactly the same instructions as for the AXB identification task and their response times were recorded from the onset of the third stimulus.

##### 4.1.2. Participants

Fourteen French participants from Paris V University, aged 19–32 years, were run. None of them had had experience with Mandarin or any other tone language and all reported normal hearing and speaking. They participated in the experiment either voluntarily or for course credit. (No Mandarin listener participated in this test: Chinese data had already been collected in Experiment 1 for the same stimuli and could be used for comparison.)

#### 4.2. Results and discussion

Results pooled across participants and syllables are shown in Figs. 8A–C. There is no sign of better discrimination around mid-continuum near the category boundary region. The reaction time data on this task only indicate that RTs are generally inversely correlated with correct discrimination level. This held true for the t2–t4 and t3–t4 continua ( $r(10) = -0.768$ ,  $p = 0.0035$ , and  $r(10) = -0.629$ ,  $p = 0.029$ , respectively), though not for the t1–t2 continuum,  $r(10) = -0.08$ . A factor that might account for the continuum differences in discrimination performance could be the psychophysical difference among the contours. Recall that the interpolation used to produce the tone continua was designed so as to have constant  $\log f_0$  as well as decibel intensity differences at each point of time between adjacent step contours. The fact that the discrimination performance of French participants is essentially “flat,” that is, does not vary in a consistent way along continua, therefore suggests that their performance is mostly determined by psychophysical factors rather than being biased by linguistic factors at any level.

On a more detailed examination, the performance profile was rather flat for the t2–t4 continuum, with a weak trend for better performance toward the tone 4 end of the continuum,  $r(10) = 0.58$ ,  $p = 0.05$ ; for the t1–t2 continuum, there was a clearer trend for a better

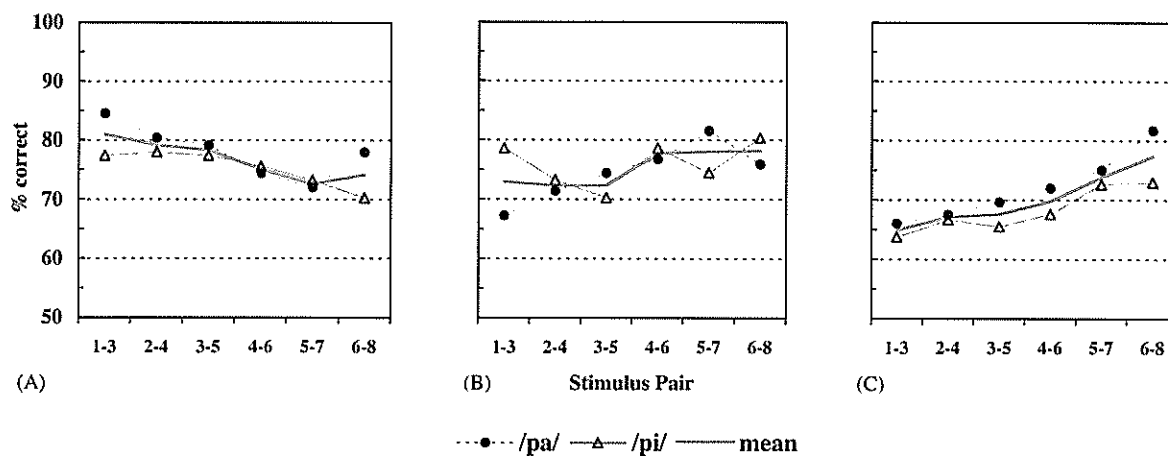


Fig. 8. (A–C). Two-step discrimination performance of French subjects in the three continua: (A) in t1–t2, (B) in t2–t4, and (C) in t3–t4.

discrimination toward the tone 1 end of the continuum,  $r(10) = -0.75$ ,  $p = 0.0052$ . A similar trend was reported by Wang (1976) for the American English subjects whom he considered to be responding in a “psychophysically motivated way,” and a subgroup of Chinese subjects who had received intensive testing in psychoacoustic experiments. The performance of the latter subjects was reported as a mix between “linguistic” and “psychophysical.” Wang argued that on a psychophysical basis, it should be easier to distinguish a level contour (such as tone 1) from a rising contour than to distinguish two differently rising contours. On the other hand, for the t3–t4 continuum, our French participants showed better performance toward tone 4, that is, the steepest end of the continuum,  $r(10) = -0.85$ ,  $p = 0.0004$ . Was this because, still on a psychophysical basis, a steep  $f_0$  and intensity drop stand out among less markedly falling contours? Whatever the correct explanations for these consistent trends for the t3–t4 continuum as well as for the t1–t2 continuum, they likely reflect a psychophysical rather than a linguistic bias. This is very unlike the Chinese participants’ discrimination performance reported above. These data thus seem to point to noncategorical processing of Taiwanese tones by French, as compared to Mandarin listeners.

Indeed, the critical issue addressed in this experiment is how the overall discrimination performance of French listeners compares with Mandarin listeners’ performance. We therefore conducted an analysis of variance which included the Taiwanese and the French discrimination data, with Language group as the between-subject variable. Because preliminary analyses conducted separately on the Taiwanese and French data showed that the Syllable type variable had no significant effect (all  $F_s < 1$ ), we pooled the data across syllable types. The within-subject variables were thus Continuum (t1–t2, t2–t4, and t3–t4) and Pair (six pairs). French participants showed significantly lower mean discrimination performance (mean 74%) than Taiwanese participants (mean 88%),  $F(1, 27) = 11.96$ ,  $p = 0.0019$ . There was a significant Language group  $\times$  Pair interaction,  $F(5, 135) = 12.10$ ,  $p < 0.00001$ , reflecting the fact that discrimination level did not significantly vary across continuum pairs for French participants,  $F(5, 65) = 1.73$ ,  $p = 0.14$ , whereas it consistently varied for Taiwanese participants,  $F(5, 70) = 18.89$ ,  $p < 0.00001$ , with a

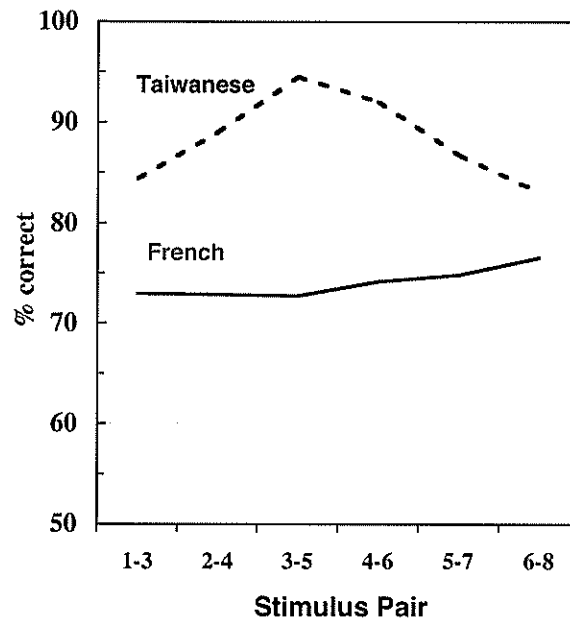


Fig. 9. Discrimination performance of French and Taiwanese participants pooled across continua (Experiments 1 vs. 3).

maximum level at mid-continuum.<sup>3</sup> The contrast between Taiwanese and French performance is illustrated in Fig. 9.

These results thus again support the view that French listeners, although showing a nonnegligible sensitivity to tone contour variations, do not process them linguistically or categorize them contrastively. Both the identification and the discrimination data suggest that their judgments of tonal contour differences and similarities are primarily motivated by perceptual factors of a general psychophysical nature. In contrast, Taiwanese listeners' judgments are clearly biased by the phonological value of tone contours.

## 5. General discussion

The results reported in this study converge to support the view that tones are perceived in a quasi-categorical way by listeners of Taiwan Mandarin Chinese, whereas they instead seem to be

<sup>3</sup>One-step AXB discrimination tests were also run with 12 additional French and 14 additional Taiwanese participants. The overall level of performance drastically dropped to an average 62.5% and 62.6% correct discrimination for French and Taiwanese, respectively. However, although French and Taiwanese participants clearly did not differ in their overall performance, the data otherwise patterned in the same way as in the two-step discrimination tests. An analysis of variance revealed a significant Language group  $\times$  Pair interaction,  $F(6, 144) = 3.04$ ,  $p = 0.0081$ , again reflecting the fact that discrimination level did not significantly vary across continuum pairs for French participants,  $F(6, 66) = 1.32$ ,  $p = 0.26$ , whereas it consistently varied for Taiwanese participants,  $F(6, 78) = 3.52$ ,  $p < 0.004$ , with higher levels at mid-continuum than toward continuum endpoints.

perceived in a more psychophysical way by listeners of French. Given the high degree of similarity between Taiwan and mainland China varieties of Mandarin, these results are likely to extend to mainland China Mandarin and, probably, to other similar tone languages.

The qualification “quasi-categorical” is intended to suggest that there is a gradient of categoricity in the perception of the various types of linguistic speech elements, as was long ago suggested by Liberman, Harris, Hoffman, and Griffith (1957) (also see Schouten & van Hessen, 1992, for a distinction of categoricity between consonants and vowels). In this study, we adopted a rather liberal view of “categorical perception.” It might be useful, however, to return briefly to this issue. A stringent definition of categorical perception requires an optimal fit between observed discrimination performance and performance predicted from identification, reflecting the strong claim that discrimination between two sounds is uniquely determined by the probability that they are labeled differently (Schouten & van Hessen, 1992). Clearly, in many cases, such an ideal fit is not obtained (Pisoni & Lazarus, 1975; Pisoni & Tash, 1974; Wood, 1976). Observed discrimination is almost always better than predicted discrimination, and this is largely due to the fact that within-category discrimination level is higher than the predicted chance level. Moreover, within-category discrimination is all the more possible when categories are loosely defined and overlap in terms of physical (and hence, psychophysical) properties. This is precisely the case with vowels, whose categories cover a large range of variation around typical vowels, as the work on category prototypes indicated a decade ago (Kuhl, 1991; Polka, 1995). Hence, there exists a theoretical debate about whether the emphasis should be put on category prototypes or on category boundaries.

Our work does not claim to settle this matter. We argue, simply, that linguistic categories do exist and must be specified in native listeners’ perception, however sharply the categories may be defined and however easily within-category differences may be perceived. In our view, boundaries are not merely “no man’s land” regions between categories: As Wood (1976) put it, “category boundary effects” cannot be explained entirely by psychophysical responses. In other words, the increased sensitivity to differences between the members of the pairs that straddle category boundaries at least partially reflects phonetic coding into linguistic categories that may, on occasion, be language-specific. A convincing way to show this is to use cross-linguistic data, as in recent studies showing how phonetic categories build up with linguistic experience (Guenther, 2002; Iverson, 2002). We followed the cross-linguistic approach in comparing listeners of Mandarin Chinese and listeners of French. The French listeners showed no increased sensitivity near category boundaries for the Mandarin Chinese tone contrasts used, whereas Mandarin listeners did show such increased sensitivity. We therefore must conclude that there is an effect of tone categories on the performance of native speakers of Mandarin Chinese. Our data show, for that matter, that tones are perceived more akin to vowels than to plosive consonants, in that within-category discrimination performance is generally high for both tones and vowels.

To summarize, the crucial argument supporting categorical perception of tones is the cross-linguistic difference observed between Mandarin and French listeners. Mandarin listeners outperform French listeners in both identification accuracy and between-category discrimination performance. French listeners’ discrimination performance is not biased by tone categories and is instead determined, *prima facie*, by psychophysical factors. In other words, in the perception tasks we used, the listeners’ performance is best explained by their native language phonology: If tones

are part of the phonological system, they are perceived as phonemic categories; otherwise, they are perceived as nonlinguistic melodic variations.

The present study thus can be viewed as yet another illustration of the native language bias in hearing the sounds of a foreign language. Here, the demonstration bears on the difficulty of perceiving non-native contrasts in the suprasegmental domain, namely, contrasts of syllabic tones. The literature in this field is quite scarce, and the most relevant study that addressed a similar issue was conducted in the late 1970s (Wang, 1976). However, because that study had serious limitations in terms of the number of subjects tested, as well as in the single tone continuum used, it has remained a matter of debate until now whether tones, like segments, are categorized linguistically by native speakers of tone language but not by speakers of nontonal languages. The present study went beyond these two limitations. In addition, due to the significant improvement in speech signal transformation techniques—in particular, pitch and time scaling techniques—we were able to produce natural-sounding, multi-dimensionally varying stimuli, which indeed was crucial here to minimize the likelihood that at least the native listeners perceive the stimuli in a nonspeech listening mode. In effect, in studies of segmental contrast using stimulus continua, it is relatively easy to maintain constant and natural suprasegmental characteristics while varying segmental ones in a controlled way. The difficulty with suprasegmental contrasts is in maintaining constant and natural segmental characteristics. We believe this result was achieved here. Thus, we are in a better position to claim that the cross-linguistic differences in tone perception we found are reliable and reflect unbiased tendencies determined by native language.

The response time data in Experiment 2 also suggest a qualitative difference in the identification of tones by Mandarin vs. French listeners. Taiwanese participants were much faster than French listeners (by about 300 ms) at identifying tones in which X was *physically identical* to A or B trials, that is, at matching tones which unambiguously belong to the same category at the endpoints of continua. This substantial advantage for the native listeners of Mandarin Chinese suggests a very differently timed categorization of tones for Mandarin vs. French listeners. This is not unrelated to the issue mentioned in the Introduction of how tone information is used in on-line processing. One view is that tone information requires time to be “computed” and might not be used on-line at all (Cutler & Chen, 1997), unless it is made predictable in some way (Ye & Connine, 1999). However, it does not seem far-fetched that there may be early cues to tone identity in both amplitude and pitch contour initial portions (see Figs. 1 and 2). Indeed, this is supported by a recent gating study by Lee (2000) who found that the “tone identification point” occurred after only 10–15% of the tone contour duration. In a previous study, Whalen and Xu (1992) showed, in a different way, that short portions of tone contours were sufficient to identify tones. Our data are also in line with the view that tone identification is faster than expected on the logical basis of whole contour availability.

So, French listeners are slower and less accurate than speakers of Mandarin Chinese on tone categorization. But are they “deaf” to tone differences? Recall that we started with the assumption that French, in particular, was a good language candidate to test “deafness” to prosodic variations because it is a nonstressed language. Indeed, Dupoux et al. (1997) have convincingly demonstrated that French listeners are “deaf” to stress contrasts in Spanish (e.g., they cannot discriminate Spanish *bébe* and *bebé*). However, French listeners’ ability to discriminate variations of tone contours is not so bad. French listeners are not deaf to tones,

although their discrimination performance is poorer than that of Mandarin listeners. We may speculate on the reasons why this is the case. One possibility is that the “not so bad” performance of French listeners reflects sensitivity to intonation contours as a means to convey emotions, nuances of opinion, attitude, and more generally both linguistic and nonlinguistic information. Such sensitivity, which is most certainly shared universally, might be enhanced in French listeners because French prosody shows little to no constraint by lexical accentuation and stress patterns (as English does). Expressive intonation in French is freed, as it were, from word-level structural constraints. Following this speculative interpretation, French listeners might outperform, for example, English listeners in perceiving tone variations. This, of course, remains an empirical issue.

From a theoretical vantage point, we might want to understand French listeners’ performance with reference to cross-linguistic models of speech perception. That French listeners are not deaf to tones is reminiscent of the finding that some contrasts, which are phonemic in one language and not in another, are not necessarily very difficult to perceive by the speakers of that other language. This has been shown, for example, in the work of Best and colleagues (Best, 1995; Best, McRoberts, & Sithole, 1988) for the Zulu click contrasts (e.g., dental vs. lateral clicks). Clicks are utterly foreign, as speech sounds, to Americans. Yet native speakers of American English *do* discriminate the contrasts reasonably well. Zulu clicks seem to be heard by Americans as a nonlinguistic sound track superimposed on a linguistic sequence of speech segments. Clearly, the classic “phonological filter” metaphor (Trubetzkoy, 1939; Polivanov, 1931) does not apply here: American listeners can discriminate click contrasts insofar as they hear them as different nonlinguistic sounds (e.g., snapping one’s fingers vs. uncorking a bottle of wine). The explanation offered by Best’s PAM (Perceptual Assimilation Model) is that clicks are not assimilable (NA) to any English sounds and are thus perceived and discriminated without calling upon language-specific phonetic perception that could assimilate them to English segments (for recent formulations of the PAM model, see Best, 1995; Best, McRoberts, & Goodell, 2001). This general idea might apply to tones as perceived by French listeners, although PAM has been designed and applied thus far only to the segmental aspects of cross-language speech perception, not the prosodic aspects. Yet, assuming that PAM can be extended to suprasegmental tiers, it would follow that French listeners have little or only moderate trouble at discriminating Mandarin tone contrasts inasmuch as they are perceived as NA contrasts. For example, French listeners can hear the difference between /ta<sub>2</sub>/ (“a dozen”) and /ta<sub>4</sub>/ (“big”). The acoustic difference is salient enough. Should we thus consider Mandarin tone contrasts as suprasegmental NA contrasts for French listeners?

The problem here is that the acoustic correlates of tones,  $f_0$  and intensity contour, are used in French, just as in any language, at the sentential intonation level. Tone contours thus are not completely irrelevant to a French ear with respect to their putative linguistic value. In that sense, we cannot consider tone contrasts to be NA for listeners of French (or any other language). But if we still consider the extension of PAM to prosodic specifications, another possible analogy can be made with what PAM would call “uncommitted” or “uncategorized” phonetic space (cf. Best, 1995). Tone contours indeed are prosodic aspects of speech that can be heard by French listeners as produced by a human voice, but are not analyzed as basic prosodic units bearing contrastive linguistic significance. By reference to the PAM framework, then, it is tempting to label tone contrasts as UU (uncategorized–uncategorized). This is a situation where PAM would predict

discrimination performances ranging from fair to good depending on the perceived salience of the phonetic (i.e., intonational) differences involved.

Perceived differences and similarities might be nonlanguage-specific in this case, or they could be evocative of language-specific intonation patterns. Tone 4, for example, could suggest a forbidding intonation to a French listener, tone 2 a stunned intonation, and so on (see Carton, 1974, pp. 91–98, for examples of such intonational values). However, these potentially meaningful traits are loosely defined and can apply to any word or sentence, whatever its length. Moreover, there exists a continuum of plausible intonations “in between” various intonation nuances so that a French listener does not have to classify any given intonation within a finite set of contrastive categories. In contrast, the linguistic function of tones in Mandarin is the same as that of phonemes. There are exactly four such phonemes—or “tonemes”—in Mandarin Chinese. They must be processed by Mandarin listeners with sufficient accuracy to be correctly identified, just like consonants and vowels. This qualitative difference in linguistic function justifies that we consider Mandarin tones as similar to phonemes for Mandarin listeners but not for French listeners. In that sense, the PAM model makes the correct prediction that French listeners can discriminate tones rather easily: They are by no means “deaf” to tonal variations. But they fail to perceive tones along the lines of a well-defined and finite set of contrastive linguistic categories, in sharp distinction to Mandarin Chinese listeners.

### Acknowledgements

This work benefited from a “Cognitique” (French Ministère de la Recherche) grant to the first author (LACO 1) and from a NSC (National Scientific Council) grant to the second author. A preliminary report bearing on the first Experiment has been published in Chinese in *Tsing Hua Journal of Chinese Studies, New Series XXX, vol. 1* (March 2000). We are thankful to all the Taiwanese and French subjects who participated in the time-consuming and attention-demanding experiments that we carried out. We also are indebted to Arthur Abramson, Jacqueline Vaissière, Doug Whalen, Laurent Demany, Juan Segui, and three anonymous reviewers who largely contributed to the improvement of the original manuscript.

### Appendix

Cumulated frequencies of the morphemes that are homophonic to the endpoint stimuli (log frequency scale). For each endpoint, the most frequent morpheme is shown in parentheses, together with an approximate English gloss.

Syllable	Tone			
	tone 1	tone 2	tone 3	tone 4
/pa/	6.70 (八 “eight”)	4.65 (拔 “pull”)	8.42 (把 “take”)	6.32 (爸 “daddy”)
/pi/	4.65 (逼 “oppress”)	5.09 (鼻 “nose”)	6.88 (比 “compare”)	6.27 (必 “must”)
/kwo/	4.35 (锅 “pan”)	7.64 (国 “country”)	5.83 (果 “fruit”)	8.03 (过 “past”)



## References

- Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, 21, 319–325.
- Abramson, A. (1979). The noncategorical perception of tone categories in Thai. In B. Lindblom, & S. Ohman (Eds.), *Frontiers of speech communication* (pp. 127–134). London: Academic Press.
- Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange, & J. Jenkins (Eds.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 45–60.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system. *Journal of the Acoustical Society of America*, 109, 775–794.
- Best, C. T., Morongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics*, 29, 191–211.
- Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20, 305–330.
- Carton, F. (1974). *Introduction à la phonétique du français*. Paris: Bordas.
- Chan, S. W., Chuang, C.-K., & Wang, W. S.-Y. (1975). Cross-linguistic study of categorical perception for lexical tone. *Journal of the Acoustical Society of America*, 58, S119.
- Chang, Y.-C., & Hallé, P. (2000). Taiwan Huayu shengdiao fanchou ganzhi [Categorical perception of Taiwan Mandarin tones]. *Tsing Hua Journal of Chinese Studies, new series XXX, 1*, 51–65.
- Chao, Y. R. (1948). *Mandarin primer*. Cambridge: Harvard University Press.
- Chao, Y.-R. (1968). *A grammar of spoken chinese*. Berkeley: University of California Press.
- Connell, B., Hogan, J., & Rozsypal, A. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics*, 11, 337–351.
- Coster, D. C., & Kratochvil, P. (1984). Tone and stress discrimination in normal Peking dialect speech. In B. Hong (Ed.), *New papers in Chinese linguistics* (pp. 119–132). Canberra: Australian National University Press.
- Cutler, A. (1986). *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, 29, 201–220.
- Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, 59, 165–179.
- Cutler, A., & van Donselaar, W. (2001). *Voornaam* is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, 44, 171–195.
- Cutler, A., & Otake, T. (1999). Pitch accent in spoken word recognition in Japanese. *Journal of the Acoustical Society of America*, 105, 1977–1988.
- Di Cristo, A. (1998). Intonation in French. In D. Hirst, & A. di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 195–218). Cambridge: Cambridge University Press.
- Dupoux, E., Pallier, C., Sebastian-Gallés, N., & Mehler, J. (1997). A destressing “deafness” in French? *Journal of Memory and Language*, 36, 406–421.
- Fon, J., & Chiang, W.-Y. (1999). What does Chao have to say about tones? A case study of Taiwan Mandarin. *Journal of Chinese Linguistics*, 27, 15–37.
- Fox, R., & Qi, Y.-Y. (1990). Context effects in the perception of lexical tone. *Journal of Chinese Linguistics*, 18, 261–283.
- Fox, R., & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics*, 13, 69–90.
- Gandour, J., & Dardarananda, R. (1983). Identification of tonal contrasts in Thai aphasic patients. *Brain and Language*, 18, 98–114.
- Gandour, J., & Harshman, R. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, 21, 1–33.

- Gårding, E., Kratochvil, P., Svantesson, J.-O., & Zhang, J.-L. (1986). Tone 4 and Tone 3 discrimination in Modern Standard Chinese. *Language and Speech*, 29, 281–293.
- Gandour, J., Petty, S., & Dardarananda, R. (1988). Perception and production of tone in aphasia. *Brain and Language*, 35, 201–240.
- Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, D. (2000). A cross-linguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, 12, 207–222.
- Gandour, J., Wong, D., & Hutchins, D. (1998). Pitch processing in the human brain is influenced by language experience. *Neuroreport*, 9, 2115–2119.
- Garner, W. R. (1970). The stimulus in information processing. *American Psychologist*, 25, 350–358.
- Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45, 133–159.
- Guenther, F. H. (2002). Effects of category learning on auditory perception and cortical maps. *Journal of the Acoustical Society of America*, 111, 2383 [abstract].
- Hallé, P. (1994). Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese. *Language and Speech*, 37, 103–124.
- Hallé, P., Best, C., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, 27, 281–306.
- Ho, A. (1976). The acoustic variation of Mandarin tones. *Phonetica*, 33, 353–367.
- Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, 30, 129–148.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge, UK: Cambridge University Press.
- Iverson, P. (2002). Perceptual interference effects on phonetic categorization by second language learners and cochlear implant patients. *Journal of the Acoustical Society of America*, 111, 2383 [abstract].
- Klein, D., Zatorre, R., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage*, 13, 646–653.
- Kolinsky, R. (1998). Spoken word recognition: A stage processing approach to language differences. *European Journal of Cognitive Psychology*, 10, 1–40.
- Kong, Q.-M. (1987). Influence of tones upon vowel duration in Cantonese. *Language and Speech*, 30, 387–399.
- Kratochvil, P. (1968). *The Chinese language today*. London: Hutchinson University Library.
- Kratochvil, P. (1985). Variable norms of tones in Beijing prosody. *Cahiers de Linguistique Asie Orientale*, 14, 153–174.
- Kratochvil, P. (1987). The case of the third tone. In: Ma (Ed.), *Wang Li memorial Volumes* (pp. 253–276). Hongkong: The Chinese Language Society of Hongkong.
- Kratochvil, P. (1998). Intonation in Beijing Chinese. In D. Hirst, & A. di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 417–434). Cambridge: Cambridge University Press.
- Kuhl, P. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93–107.
- Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics*, 11, 373–382.
- Lee, C.-Y. (2000). *Lexical tone in spoken word recognition: A view from Mandarin Chinese*. Unpublished doctoral dissertation, Brown University. Providence, RI.
- Lee, L., & Nusbaum, H. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception and Psychophysics*, 53, 157–165.
- Lee, Y.-S., Vakock, D., & Wurm, L. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 125, 527–542.
- Liberman, A., Haris, K., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.
- Massaro, D. (1987). Categorical partition: A fuzzy logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 254–286). Cambridge, MA: Cambridge University Press.
- Massaro, D., & Cohen, M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, 2, 15–35.
- Moulines, E., & Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Communication*, 16, 175–205.

- Pisoni, D., & Lazarus, J. (1975). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 55, 328–333.
- Pisoni, D., & Tash, J. (1974). Reaction times to comparisons within and across phonetic boundaries. *Perception and Psychophysics*, 15, 285–290.
- Polivanov, E. (1931). La perception des sons d'une langue étrangère [Perception of the sounds of a non-native language]. *Travaux du Cercle Linguistique de Prague*, 4, 79–96.
- Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America*, 97, 1286–1296.
- Repp, B., & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics*, 18, 481–495.
- Rossi, M. (1980). Le français, langue sans accent? [French: A non-stressed language?]. In I. Fónagy, & P. Léon (Eds.), *L'accent en français contemporain [Stress in modern French]* (pp. 13–51). Paris: Didier.
- Schouten, M., & van Hessen, A. (1992). Modeling phoneme perception. I: Categorical perception. *Journal of the Acoustical Society of America*, 92, 1841–1855.
- Semal, S., Demany, L., Ueda, K., & Hallé, P. (1996). Speech vs. nonspeech in pitch memory. *Journal of the Acoustical Society of America*, 100, 1132–1140.
- Soto-Faraco, S., Sebastian-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, 45, 412–432.
- Stagray, J., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and nontone language. *Journal of Chinese Linguistics*, 21, 143–163.
- Stevens, K. (1966). On the relations between speech movements and speech perception. Paper presented at the meeting of the 18th International Congress of Psychology, Moscow, August 1966.
- Troubetzkoy, N. S. (1939). Grundzüge der Phonologie [Principles of Phonology]. *Travaux du Cercle Linguistique de Prague*, 7, 1–271.
- Vaissière, J. (1991). Rhythm, accentuation, and final lengthening in French. In J. Sundberg, & R. Carlson (Eds.), *Music, language, speech, and brain* (pp. 108–120). New York: Macmillan Press.
- Van Lancker, D., & Fromkin, V. (1973). Hemispheric specialization for pitch and “tone”: Evidence from Thai. *Journal of Phonetics*, 1, 101–109.
- Wang, H., Chang, B., Li, Y., Lin, L., Lin, J., Sun, Y., Wang, Z., Yu, Y., Zhang, J., & Li, D. (1986). *Frequency dictionary of contemporary Chinese*. Beijing: Beijing Language Institute Press.
- Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences*, 28, 61–72.
- Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language*, 78, 332–348.
- Whalen, D., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25–47.
- Wood, C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America*, 60, 1381–1389.
- Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, 95, 2240–2253.
- Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, 55, 179–203.
- Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33, 319–337.
- Yang, Y.-F. (1991). Ear differences in distinguishing consonant features and lexical tones. *Acta Psychologica Sinica*, 2, 131–137.
- Ye, Y., & Connine, C. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, 14, 609–630.