Journal of Phonetics 40 (2012) 672–688

Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Cross-linguistic perception of Itunyoso Trique tone

Christian T. DiCanio*

Laboratoire Dynamique du Langage, CNRS/Université Lumière Lyon-2, France

ARTICLE INFO

Article history:
Received 22 March 2011
Received in revised form
2 May 2012
Accepted 15 May 2012
Available online 15 June 2012

ABSTRACT

Recent findings have argued in favor of the categorical perception of tonal contrasts in Taiwanese Mandarin and Standard Mandarin (Hallé, Chang, & Best, 2004; Xu, Gandour, & Francis, 2006), and most recently in Mandarin and Cantonese (Peng et al., 2010). Findings in favor of the categorical perception of tone emerge most clearly from cross-linguistic work on speech perception. The current study continues this line of research by investigating the categorical perception of Itunyoso Trique tone among Trique and French listeners. Tonal stimuli were presented to listeners in an AXB discrimination task (2AFC) and an AXB identification task (2AFC), closely following methods used in Hallé et al. (2004). Evidence for a listener sensitivity to tonal categories was found for Trique listeners in their discrimination performance, but this pattern did not correspond to the identification performance. Overall, French speakers performed better overall at tone discrimination than Trique listeners, who largely ignored within-category phonetic differences. Both Trique and French listeners were found to be sensitive to psychoacoustic differences between stimuli, though French speakers relied more heavily on such differences. The findings here argue for the importance of both phonetic and auditory memory for the perception of Trique lexical tone.

1. Introduction

In early work on speech perception, researchers observed a close correspondence between the discriminability of a phonological contrast and its identification (Liberman, Haris, Hoffman, & Griffith, 1957). This phenomenon, called *categorical perception*, was considered a special mode of auditory perception that applied only to speech phenomena. Subsequent research largely discredited the assertion that categorical perception was a phenomenon specific to speech (Harnad, 1987). For instance, sharper categorical boundary effects were observed for auditory contrasts with more abrupt rise-times, such as the discrimination between "plucked" and "bowed" notes on a musical continuum (Cutting & Rosner, 1976; Macmillan, 1987). Furthermore, the identification boundaries between many phonological contrasts have been observed to straddle regions of greater psychoacoustic sensitivity, such as natural boundaries in temporal order discrimination (Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pastore, 1987). If these psychoacoustic factors account for the observed relationship between identification and discrimination, then surely categorical perception is domain-general.[1]

However, these critiques do not imply that psychoacoustic differences explain all categorization in speech. Modern research in speech perception is concerned with how psychoacoustic sensitivities, memory, phonetic cues, and language biases all interact in speech categorization. Generally speaking, consonantal contrasts will have abrupt perceptual boundaries where a peak in the discrimination function between categories corresponds closely with an abrupt change in category identification (Francis, Ciocca, & Ng, 2003; Harnad, 1987; Repp, 1984). Vowel contrasts are perceived with less abrupt category boundaries, owing to their distinct representation in auditory short-term memory (Fry, Abramson, Eimas, & Liberman, 1962; Pisoni, 1973; Rosen & Howell, 1987). Yet, with respect to tonal contrasts, the findings have been mixed. Some researchers have found tones to have abrupt, but fuzzier identification boundaries like vowels (Avelino, 2003; Chan, Chuang, & Wang, 1975; Hallé et al., 2004; Stagray & Downs, 1993; Xu et al., 2006) while others have argued that tonal contrasts are perceived more continuously (Abramson, 1979; Francis et al., 2003).

In a recent study examining the perception of Taiwanese Mandarin tone with Taiwanese and French listeners, Hallé et al. (2004) found that native listeners' discrimination peaks corresponded closely with identification crossover locations. The identification functions for Taiwanese listeners were not as

* Tel.: +203 865 6163.
E-mail address: cdicanio@gmail.com

[1] Additionally, work on animal perception has found similar identification boundaries in chinchillas and macaques for a VOT (voiced-voiceless) contrast in English to those observed in humans (Kuhl & Miller, 1975; Kuhl & Padden, 1982; Pastore, 1987). If the sharpness of the identification slope is a characteristic of categorical perception, then such evidence argues for its domain generality.

0095-4470/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2012.05.003

abrupt as those typically observed for consonantal contrasts, but they showed a significantly steeper slope between category endpoints on the continuum. For Taiwanese Mandarin listeners, there was greater sensitivity to pitch and intensity differences when they corresponded to categorical boundaries between tones. These effects were not observed for French listeners. Hallé et al. argue that tone is perceived *quasi-categorically* in Taiwanese Mandarin. Given that the perception of phonological contrasts lies on a continuum of categoricity largely dependent on the experimental task and its specific demands on phonetic and auditory memory (Gerrits & Schouten, 2004), the term '*quasi-categorical*' is a bit of a misnomer. Nevertheless, these results are particularly interesting because categorization by Taiwanese listeners is not explained by the experimental condition nor the psychoacoustic differences between the stimuli. The study involved the cross-linguistic perception of a contrast that does not exist in French. If psychoacoustic differences between the stimuli accounted for the discriminability of tonal contrasts, one would expect similar behavior for French and Taiwanese Mandarin listeners. Cross-linguistic work on non-native contrasts provides a way to control for both auditory sensitivities and experimental procedure while investigating evidence of categorization.

As approximately 42% of the world's languages have tonal contrasts (Maddieson, 2010), understanding how they are perceived and categorized is a crucial part of understanding human cognitive capacities. Yet, the perception of tone also offers a unique window into how cross-linguistic speech perception works. All languages have segments which are used to create phonological contrasts, but not all languages have tone. A tone language from the perspective of a non-tone language speaker (and vice-versa) represents a completely different organization of prosodic structure, a larger structural difference than one which is merely segmental. While tone is common in the world's languages, tone perception has been studied only within a small group of languages to date. Most work has exclusively focused on East and Southeast Asian languages (Brunelle, 2009; Francis et al., 2003; Gandour & Harshman, 1978; Hallé et al., 2004; Lee, Vakoch, & Wurm, 1996; Peng, 1997; Xu, 1994; Xu & Xu, 2003, among others). Exhaustively, the exceptions are Connell (2000) on Mambila, Hombert (1975, 1976), and Lavelle (1974) on Yoruba, Avelino (2003) on Zapotec, and, more recently, Frazier (2009) on Yucatec Maya.

The current paper addresses the categorical perception of tone in Itunyoso Trique, an Oto-Manguean language spoken in Mexico (DiCanio 2008, 2010b, 2012a, 2012b). The results from two experiments are presented, which tested how well speakers of Itunyoso Trique and French perceive Trique tonal contrasts. The primary objectives of this study are to examine the differences in tonal discrimination and identification between Trique and French listeners and to examine if listeners are sensitive to tonal categories in tone perception. The results show evidence for greater *between*-category than within-category discrimination for Trique listeners and more accurate *within*-category discrimination among French listeners than among Trique listeners. Overall, French listeners were more accurate discriminating the tonal contrasts, but did not show any evidence of categorical perception in discrimination. The identification functions varied between languages, but no relationship between identification and discrimination performance was found. In general, among tonal contrasts which were harder to discriminate, there was less of a relationship between identification and discrimination performance.

This paper also contributes to a growing literature investigating language biases of non-native speech contrasts (Best & Strange, 1992; Cho & McQueen, 2006; Hallé, Best, & Levitt, 1999, 2004; So, 2006). Previous work has argued that speakers of tonal languages have an advantage over speakers of non-tonal languages in tonal discrimination tasks due to their native language experience with pitch as a linguistic cue (Burnham et al., 1996; Lee et al., 1996; Hallé et al., 2004; So, 2006). The results here argue against this showing that French listeners have an advantage over Trique listeners in tonal discrimination. Trique listeners are less accurate overall in tone discrimination than French listeners, but largely seem to ignore intra-categorical differences in pitch, similar to findings in Stagray and Downs (1993).

This paper is organized as follows. In Section 1.2, I provide background information on the tonal system of Itunyoso Trique. In Section 1.1, I discuss the literature on tone perception. In Section 2, I discuss the methods and results from an AXB tone discrimination task with Trique and French listeners. In Section 3, I discuss the methods and results from an AXB tone identification task with Trique and French listeners. In Section 4, I provide a general discussion of the findings from this series of experiments. I conclude the paper in Section 5.

1.1. *The perception of lexical tone*

Research on tone perception can be divided into two areas typical of work on speech perception: categorization studies and studies on perceptual cues. In this section, I provide a summary of the literature on these topics.

1.1.1. *Tonal categorization*

Repp (1984) notes that categorical perception shows four main characteristics. First, one observes an abrupt change in the identification responses (labeling probabilities) between two stimuli when they are presented along a regularly spaced continuum. Second, one observes a peak in discriminability which corresponds to this abrupt change in identification. Third, discrimination is at or near chance level within a category. Fourth, the discrimination function is perfectly predictable from the labeling probabilities.

Typically, one does not observe this idealized pattern of categorical perception. Discrimination performance is usually better than predicted from labeling probabilities (Francis et al., 2003), controverting the final two of these characteristics. Furthermore, the abruptness of the labeling probabilities varies continuously depending on the degree to which auditory memory is recruited for categorization (Francis et al., 2003; Repp, 1984). For instance, one observes a more gradual identification function in identification tasks with a short ISI. Since auditory memory decays more quickly than phonetic memory, listeners are able to utilize auditory memory to a greater degree in such conditions (Pisoni, 1973; Xu et al., 2006). Finally, the meaning of "abruptness" varies depending on the contrast type. What is considered an abrupt change in identification for a vowel contrast might be considered to be gradual for a consonantal contrast, since vowels are retained longer in phonetic memory (Pisoni, 1973).

Yet, once these factors are considered, how do listeners categorize tonal contrasts? Does one observe steeper slopes on the labeling function during identification, as observed for certain vowels, or gradual, continuous identification responses? The earliest studies on tone categorization concern the identification of tones by native speakers of Thai (Abramson, 1962, 1975) and Mandarin Chinese (Chuang, Hiki, Sone, & Nimura, 1972; Howie, 1972). In each of these studies, listeners were asked to identify words on the basis of their tone. Listeners identified tones with a high degree of accuracy, although certain tones (M vs. L in Thai, Tone 2 vs. Tone 3 in Mandarin) were harder to identify than others. Later work on Thai (Abramson, 1979) and Mandarin (Chan et al., 1975) explored the categorical perception of tone using a traditional paradigm, i.e. per Liberman et al. (1957). Abramson tested Thai listeners on a level tone continuum spanning the pitch range of the three level tones in the language. Thai listeners did not have discrimination peaks between tonal categories, which led Abramson to conclude that tone was not perceived categorically. Chan et al. reached a separate conclusion for Mandarin Chinese. Testing an 11 step acoustic continuum between tone 1 and tone 2, they found a discrimination peak corresponding to an identification crossover point between the tonal categories. This is classical evidence of categorical perception, which led them to conclude that tone was categorically perceived. Even though both researchers used the same paradigm, they reached separate conclusions.

Mixed findings continued to persist in later work on the categorical perception of tone. In a same-different tone discrimination task, Stagray and Downs (1993) found that Mandarin listeners were less sensitive to within-category pitch differences than English listeners were. They argued that *within*-category pitch variation was ignored by listeners because it is not relevant to categorical identification. Listeners ignored phonetic variability which was not contrastive. In a tonal identification task on Mambila, Connell (2000) tested listeners using a 10-step continuum spanning the language's four level tones. The identification boundaries between the four level tones were fairly gradual and no abrupt change in identification occurred. While Connell considered this positive evidence that pitch height alone could cause a change in identification for Mambila listeners, it was less conclusive that listeners had discrete category boundaries of the sort predicted in a categorical mode of perception. In his work on Zapotec, Avelino (2003) found that listeners performed well in tonal identification, but did not consistently discriminate between tones. A categorical effect was found between the two level tones (H vs. L) in the language, but not between the level and the contour tones. In a study on Cantonese tone perception, Francis et al. (2003) argue that psychoacoustic effects and learned category boundaries account for peaks in discriminability between tonal categories. Moreover, the authors found that identification crossover points did not predict the location of peaks in discrimination. This finding argues against the hypothesis that the discriminability of contrasts reflects long-term memory representations used in categorical identification. Instead, they argue that listeners attend to natural regions of auditory sensitivity in discrimination.

The general lack of consistent findings on the categorical perception of tone may result from the nature of the stimuli being examined. In many studies, tonal stimuli are not presented in context, but in isolation (Abramson, 1979; Avelino, 2003; Connell, 2000). It is well-known that listeners rely on contextual cues for determining the tonal space of the speaker (Francis, Ciocca, King Yu Wong, Ho Yin Leung, & Cheuk YanChu, 2006; Wong & Diehl, 2003). Without these cues, listeners may have difficulty making categorical phonological judgments. Furthermore, discrimination between level tones may be fundamentally different from discrimination between contour tones, which are self-normalizing (Francis et al., 2003). This may account for the general lack of categorical perception effects among those studies examining level tones, e.g. in Thai and Mambila. In contrast, categorical perception of tone is attested among studies examining contour tones, e.g. in Mandarin and Cantonese.

One way to control for auditory sensitivities in tonal discrimination is to examine tone perception cross-linguistically. For instance, the lack of discrimination boundary effects for French listeners in the experiments discussed in Hallé et al. (2004) suggests that peaks in sensitivity do not have a simple auditory explanation. Furthermore, in a cross-linguistic study of tone in Mandarin Chinese, Xu et al. (2006) found categorization boundary effects for Chinese listeners, but none for English listeners. This finding leads Xu et al. to argue in favor of a multistore model of categorical perception, rather than one which is solely domain-specific or domain-general. A similar type of argument in favor of parallel memory representations in categorical perception is made in Carney, Widin, and Viemeister (1977) and Huang and Johnson (2010). In both the works of Francis et al. and Xu et al., listeners of a non-tonal language were used as a comparison with native listeners of a tonal language. In doing so, one controls for psychoacoustic boundary effects which are anticipated to be identical across language groups.

1.1.2. Tonal cues

Relatively few studies have examined the relative contribution of different acoustic cues in the perception of tonal contrasts. The main perceptual cues of most tonal contrasts are F_0 height, F_0 trajectory, and F_0 slope. However, other cues may also contribute to tone perception. In a pioneering study, Gandour and Harshman (1978) investigated the ability of English, Thai, and Yoruba speakers to discriminate synthetic pitch stimuli on open syllables. They observed that speakers of tone languages attached greater perceptual weight to dynamic pitch dimensions, such as pitch direction and slope, than to static dimensions like average pitch height. Similar patterns have been observed for Taiwanese (Lin & Repp, 1989) and Thai listeners (Zsiga & Nitisaroj, 2007), suggesting that there is a general tendency for speakers of languages with contour tones to weigh tone movement more than heavily than tone height in perception.

Apart from these cues, a number of non-pitch cues are also used in tone perception. Duration is used as a cue in tone perception for Mandarin Chinese (Liu & Samuel, 2004) and for Taiwanese listeners (Lin & Repp, 1989). Voice quality (glottalization) is used as a tonal cue for listeners of Vietnamese (Brunelle, 2009) and Yucatec Maya (Frazier, 2009). Pitch differences between tones also introduce natural differences in intensity. Such differences may also be used as cues in tonal perception (Whalen & Xu, 1992).

Findings like those observed in Hallé et al. (2004) and Xu et al. (2006) demonstrate that cross-linguistic perceptual studies are a constructive paradigm for investigating categorical boundary effects for tone. Studies investigating language-bias in tone perception have revealed stark differences in discrimination sensitivity among speakers from different language backgrounds (Francis, Ciocca, Ma, & Fenn, 2008; So, 2006) or even from different dialect backgrounds (Brunelle, 2009). Within the same spirit, the current study compares the perception of French and Trique listeners on a Trique tonal contrast. As categorical perception effects were found using the methods in Hallé et al. (2004), the methods used here closely followed those used by these authors. In the first experiment, Trique and French subjects participated in an AXB discrimination task. In the second experiment, Trique and French subjects participated in an AXB identification task. This task differed from traditional identification tasks as listeners heard tone category prototypes surrounding the target stimulus. This experiment is also unique from most work in tonal perception in that all Trique subjects are illiterate in their native language, as there is no commonly used writing system (but see DiCanio & Cruz Martínez, 2010, for a recent attempt at providing one). Most speakers of tonal East and Southeast Asian languages are literate.

1.2. The phonetics of Itunyoso trique tone

Before reporting the present study, it is necessary to describe the phonetic characteristics of tone in Itunyoso Trique, an Oto-Manguean language spoken in Oaxaca, Mexico. The acoustic properties of tone in Trique are comprehensively described in DiCanio (2008). There are nine contrastive tones in Itunyoso Trique: four level tones and five contour tones. Unlike many East-Asian and Southeast Asian tonal languages, most words in Itunyoso Trique are polysyllabic. Contrastive tone occurs on every syllable in the word, but the full set of tonal contrasts may only occur on the final syllables of lexical roots. Moreover, these nine different tones are unevenly distributed in final syllables of three possible rime shapes: long vowels without a coda (/Vː/), short vowels with a voiced glottal fricative coda (/Vɦ/), and short vowels followed by a coda glottal stop (/Vʔ/). Non-final syllables are always open with a short vowel. A chart exemplifying the tonal contrasts in monosyllabic words is given in Table 1. Tonal transcription here follows Chao's tone letters, where /1/ is the lowest pitch level and /5/ is the highest.

The four level tones differ primarily in pitch height and are realized without any significant pitch contours. The falling tones /43/, /32/, and /31/ each contrast in pitch height. Tones /43/ and /32/ have a similar pitch declination (26.5 vs. 19.4 Hz, respectively). Tone /31/ is realized with a more abrupt falling contour than either tone /43/ or /32/ (33.5 Hz). The rising tones differ only in pitch height and do not differ in slope. Fig. 1 shows average pitch traces for tones in Vː contexts on monosyllabic roots for 6 speakers.

As the current paper is concerned with the perception of tones on open syllables, the phonetics of only these tones are described here. All values here come from DiCanio (2008). The level tones are not evenly spaced. At their midpoints, the difference between tones /1/ and /2/ is 11.7 Hz, while the difference for tones /2/ and /3/ is 21.3 Hz, and 18.8 Hz between tones /3/ and /4/. Moreover, tone /1/ is the shortest in duration of the level tones while tone /3/ is the longest. The contour tone /43/ begins at a higher pitch level than tone /4/ and falls to approximately the level of tone /3/, with a pitch excursion of 21.0 Hz. Tone /32/ begins at approximately the pitch level of tone /3/, but has a small pitch excursion of 14.3 Hz. Tone /31/ begins at a pitch level midway between that of tone /3/ and /2/ but terminates at a pitch level lower than tone /1/, having a large pitch excursion of 27.4 Hz. (Note that these pitch excursions are estimated not from the vowel endpoints, but from the middle 80% of the vowel's duration.) Tone /43/ is the shortest of all tones, while tone /32/ is the longest.

While tonal allophony has been shown to affect listeners' judgments of phonological similarity (Huang, 2004; Huang & Johnson, 2010), there are few such processes in Itunyoso Trique. Tonal alternations are always motivated by the morphological system and occur within the word. For instance, aspect-marking on verbs causes tonal changes on all or part of the word, e.g. /ki^{3}rə̃^{4}ʔə̃^{43}/ '*I have danced.*' vs. /ki^{2}rə̃^{2}ʔə̃^{2}/ '*I will dance.*'. Tones /3/ and /4/ contrast with tone /2/, which marks the potential aspect (see also Hollenbach (1984)). Person-marking enclitics also may cause tonal alternations depending on the tone of the root to which they attach, e.g. /ka^{3}lĩʔ^{3}/ '*problem*' vs. /ka^{3}kĩh^{5}/ '*my problem*'. Apart from these two types of alternations, there are no other processes which alter the tones on words in a sentence. (The reader is referred to DiCanio, 2010a for an in-depth discussion of Trique tonal morphophonology.) There are no regular tone sandhi processes that apply between words.

Table 1
Contrastive tones in Itunyoso trique.

Tone	Vː		Vɦ		Vʔ	
/4/	βːe^{4}	'*hair*'	jə̃h^{4}	'*dirt, wax*'		
/3/	nːe^{3}	'*plough*'	jə̃h^{3}	'*paper*'	tsiʔ^{3}	'*agave liquor*'
/2/	nːe^{2}	'*to lie*'	tah^{2}	'*delicious*'	tʃːiʔ^{2}	'*ten*'
/1/	nːe^{1}	'*naked*'	tah^{1}	'*how (Q)*'	tsiʔ^{1}	'*sweet*'
/43/	li^{43}	'*little*'	nːə̃h^{43}	'*mother (vocative)*'		
/32/	nːe^{32}	'*water*'	nːə̃h^{32}	'*cigarette*'		
/31/	nːe^{31}	'*meat*'				
/45/			joh^{45}	'*forehead*'		
/13/			joh^{13}	'*light, fast*'		

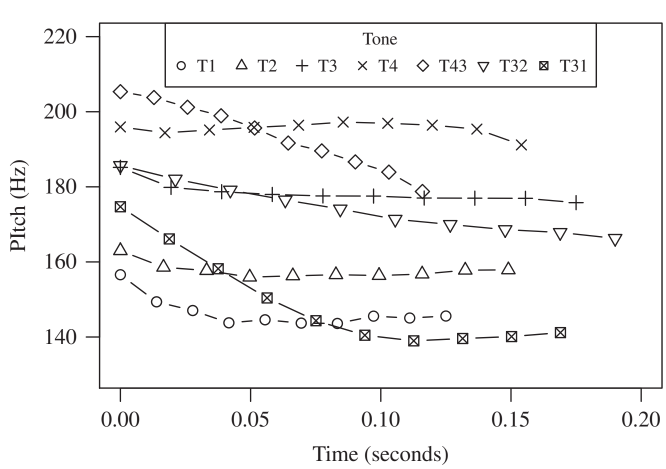

Fig. 1. Average pitch of tone in Vː contexts, from DiCanio (2008).

2. Experiment 1: tonal discrimination with Trique and French listeners

2.1. Method

2.1.1. Subjects

Eighteen Trique subjects and 20 French subjects participated in the experiment. Data from four Trique subjects were eliminated from the experiment. One Trique subject was removed due to a program error on the experimenter's computer. The other three subjects were eliminated because their responses reflected chance-level discrimination (between 50–55%). It was unclear that these subjects understood the experiment. The Trique part of the experiment took place in the town of San Martín Itunyoso. Subjects were seated in a quiet room at a desk facing the author's computer. The French part of the experiment took place in a laboratory at Université Lyon 2. The Trique portion of the experiment was administered by the author in Spanish, while the French portion was administered by the author in French.

All the subjects of both language groups were between 18 and 45 years old at the time of the experiment and none had any history of speech or hearing disorders. The average age was 24 years old for French subjects, but 26 years old for Trique subjects. Trique listeners had a larger age range (18–45 years old) than the French listeners (18–34 years old). Moreover, most of the French subjects were college-age students at least passively familiar with the expectations of experimental research. In contrast, there is little or no familiarity with research in higher education among Trique speakers. Most of the Trique subjects had received college or secondary school education, but some of the older subjects had little formal education.

Prior to the experiment, French subjects were asked if they had any musical training. Thirteen of the French subjects reported having musical experience of varying levels, but this information was not requested from the Trique subjects. All the Trique subjects were native speakers of the San Martín Itunyoso dialect of Trique and bilingual in Spanish. Bilingualism is typical of most

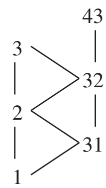

Fig. 2. Tonal comparisons.

Table 2
Tonal stimuli.

Trique word	English gloss
n:e^3	'*plough*'
n:e^2	'*to lie*'
n:e^1	'*naked*'
n:e^{32}	'*water*'
n:e^{32}	'*meat*'
ru^{4}ne^{43}	'*avocado*'
ru^{3}ne^{32}	'*bean*'

of the population in the community and typical of most American Indian languages in Mexico. However, Trique is the main language of daily and family use between Trique speakers. This is true even among younger speakers of the language. In fact, prior to their formal education, Trique children do not speak any Spanish. Spanish is used only with outsiders, in government, and in the educational system. As the French part of the experiment was carried out in France, and not the United States, most of the French subjects had some familiarity with foreign languages, typically English, Spanish, and German. However, none of the French subjects reported fluency in a language other than French.

2.1.2. Stimuli

In the first experiment, tonal stimuli were compared from open syllable words in Itunyoso Trique. There are seven contrastive tones which occur in this context (see Fig. 1). Using a two-alternative forced choice paradigm, there are 21 possible tonal comparisons.[2] For the purposes of this study, the set of comparisons was restricted to those tones which were close in the acoustic space and those for which an intermediate stimulus between categories would not span the acoustic-perceptual space of a different tone. For instance, creating resynthesized stimuli between level tones /1/ and /3/ would necessarily cross into the perceptual space of tone /2/. Comparisons like these were excluded from the study. Eight tonal category comparisons were created for an AXB discrimination task, shown in Fig. 2. Six of the seven tonal contrasts appearing in long open syllables were compared (tone level /4/ was omitted). The words used to create resynthesized stimuli are given in Table 2. Fig. 2 shows the different tonal pairs which were compared in the experiment.

For each tonal category, a representative token among the stimuli set was chosen. The duration of each of the monosyllabic tokens was neutralized to the average duration among them (655 ms). The same was done for the disyllabic words ('*avocado*' and '*bean*') (800 ms). Pitch and intensity values for each target word were extracted using a script written for Praat (Boersma & Weenink, 2009). Following the methods used in Hallé et al. (2004), six between-category stimuli were then linearly interpolated from these acoustic measures, resulting in a total of eight stimuli for each the eight tonal comparisons. Following Hallé et al. (2004), a linear scale was used for pitch resynthesis, instead of ERB (equivalent rectangular bandwidth). In all cases, resynthesis was done using the stimuli with a higher pitch value, e.g. using tone /3/ to resynthesize intermediate tokens between /3/ and /2/. Stimuli at the ends of the continua were not resynthesized, but consisted of the original time-normalized tokens.

As the current study attempts to replicate, as much as possible, the methodology in Hallé et al. (2004), pitch and intensity were resynthesized together; they were not independently manipulated. However, intensity varied slightly across tones, where falling tones had greater overall intensity than level tones (see Fig. 3). It is known that listeners of certain languages are able to rely on intensity differences alone to distinguish tonal contrasts (Liu & Samuel, 2004; Whalen & Xu, 1992), but they mainly rely on F_0 when such information is given (Abramson, 1972). Moreover, it is conceivable that such listeners are actually relying on a close correlation between absolute F_0 differences and amplitude to determine tonal identity (Whalen & Xu, 1992). While native and non-native listeners of Trique tone may use intensity information in tonal discrimination, this issue was not examined in the present study.

Pitch resynthesis was done using Praat (Boersma & Weenink, 2009) and intensity resynthesis was done using a Matlab script. An illustration of the interpolated pitch pattern is shown in Fig. 3. Following resynthesis, each word was embedded in a carrier phrase: [ka^{2}tah^2__] '*Someone will say* __'. A Trique native-speaker judged the resulting phrases with the resynthesized tokens to sound natural.

2.1.3. Procedure

An AXB discrimination paradigm was used in this experiment. All eight tonal comparisons were blocked and presented to subjects in a pseudo-random order. For each tonal comparison, subjects listened to three stimuli. The first and last stimuli in each trial were separated by two intervals along the resynthesized continuum, e.g. 1-3, 2-4, 3-5, 4-6, 5-7, 6-8. The middle test stimulus was identical to either the first or the last stimulus, e.g. 1-1-3, 1-3-3, 3-5-5, 5-5-3. These stimuli were separated by a 500 ms inter-stimulus interval (ISI). Subjects were asked to determine if the middle stimulus was more similar to the first or to the last. Each stimulus was tested eight times (twice for each presentation order: BBA, BAA, ABB, AAB), for a total of 48 trials per tonal comparison block (384 total trials). All stimuli were presented using a script written for Praat on the author's Mac iBook G4 laptop with headphones. This test phase was preceded by a practice block of 32 trials, which consisted of four trials from each of the eight tonal comparisons. French listeners used a mouse to click one of two labels on the screen, '*premier*' and '*troisième*', corresponding to the first and last token. The labels appeared in Spanish, '*primero*' and '*tercero*', for Trique listeners. The author was present during the experiment for all subjects in both language groups and explained the procedure to subjects in French or in Spanish.

2.2. Results

Four different statistical tests were used to examine the discrimination data. In all logistic models, the ANOVA(lrm) function in R (R Development Core Team, 2009) was used because it provides a test of the hypothesis that each predictor has a non-zero regression coefficient. Note that "ANOVA" here reports an analysis of deviance, not variance. The significance of each of these predictors is evaluated with a χ^2 statistic (Baayen, 2008).

[2] Where n is equivalent to the number of stimuli, $\sum_{k=1}^{n}(n-k)$ = the number of possible comparison pairs for the stimulus set. Thus, for seven tones, there are 21 (6+5+4+3+2+1) possible pairs.

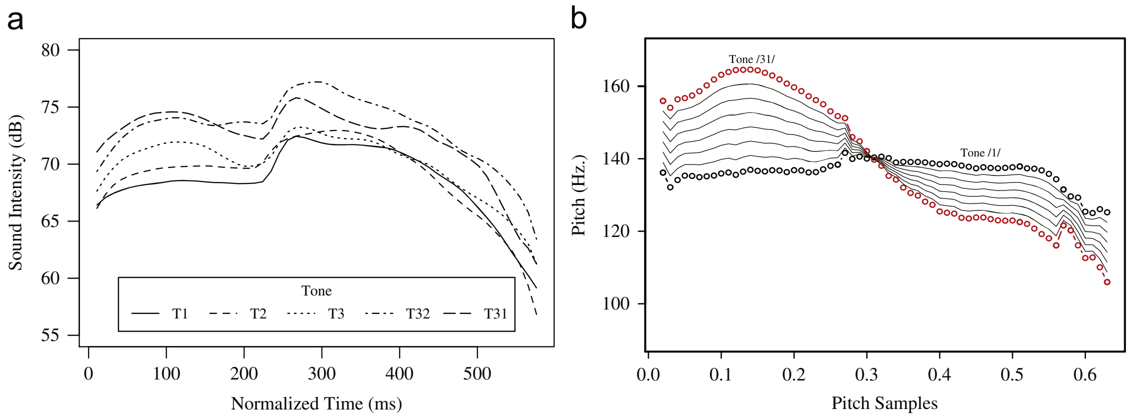

Fig. 3. Intensity differences by tone (left) and pitch resynthesis (right).

In the first test (test 1), discrimination accuracy was analyzed with a two factor logistic regression model with continuum step as a within-subjects factor, and language as a between-subjects factor. All tonal continua were grouped in the first analysis (with tonal continuum treated as another within-subjects factor), but then divided in subsequent analyses. Alpha levels for multiple comparisons were adjusted using the Bonferroni correction. This model was analyzed using ANOVA with a χ^2 test. Outliers were found in only three tonal continua: T2-T1, T2-T3, and T32-T31. They accounted for only 0.8% of the data within these blocks. Removal of outliers was completed following methods used in Baayen (2008, Section 6.2.3). After removing these outliers and re-fitting the logistic model to the data, no differences in the statistical results were found. This particular test was used to examine both general language effects and specific language $\times$ continuum step interactions for each tonal continuum.

In the second test (test 2), a separate logistic model was fit to the data where the continuum steps were divided into two groups: a *between-category* group (step comparisons 2-4, 3-5, 4-6, 5-7) or a *within-category* group (step comparisons 1-3, 6-8). This grouping and the tonal continuum were treated as within-subjects factors, while language was treated as a between subjects factor. This particular test was used to examine language effects within and across tonal categories. In the third test (test 3), listener sensitivity (d') was analyzed in a three factor ANOVA with tonal comparison and continuum step as within-subjects factors and language as a between-subjects factor. This test was the only one used to examine listener sensitivity.

In the fourth test (test 4), the peakedness of the discrimination curve was analyzed, using methods similar to those in Xu et al. (2006). A *between-category* peak in discrimination accuracy for each listener was calculated (P_{bc}) for each listener at each tonal continuum. This value was compared with the average accuracy for *within-category* discrimination, where $P_{wc}=(P_{13}+P_{68})/2$. The absolute difference between these values is a measure of peakedness in discrimination accuracy. This value was analyzed in a two factor ANOVA with tonal comparison as a within-subjects factor and language as a between-subjects factor. This test was similar to the second test (above), but offered a more precise way to examine the peak discrimination accuracy between groups (as opposed to the average).

2.2.1. Discrimination accuracy and sensitivity

There was a general effect of language on discrimination accuracy across all tonal comparisons ($\chi^2(48)=227.6$, $N=13{,}056$, $p<.001$). French listeners (78.0%) showed greater discrimination accuracy than Trique listeners (71.6%). This test reflects differences in overall discrimination accuracy (test 1). The results from test 2 showed a significant main effect of within vs. between category discrimination ($\chi^2(16)=152.8$, $N=13{,}056$, $p<.001$) and a significant interaction with language as well ($\chi^2(8)=55.2$, $N=13{,}056$, $p<.001$). These findings reflect a pattern where French listeners were much better at within-category discrimination than between-category discrimination as compared to Trique listeners.

Results from the first logistic model (test 1) show differences in discrimination accuracy between language groups for particular steps along the stimulus continuum. Fig. 4 shows the discrimination performance for Trique (T) and French (F) listeners for level tone comparisons. A significant language $\times$ continuum interaction was found for the /T3/-/T2/ comparison ($\chi^2(5)=12.1$, $N=1632$, $p<.05$) but no such effect was observed for the /T2/-/T1/ comparison ($\chi^2(5)=3.1$, $N=1632$, $p=.68$). General discrimination accuracy was low for Trique and French listeners for the /T2/-/T1/ pair. For Trique listeners, it approached chance level at the midpoint of the continuum. The low discriminability of this comparison is discussed in Section 2.3.

Fig. 5 shows the discrimination performance for tonal pairs /T1/-/T31/ and /T2/-/T31/. A significant language x continuum interaction was found for both the /T1/-/T31/ pair ($\chi^2(5)=15.5$, $N=1632$, $p<.01$) and the /T2/-/T31/ pair ($\chi^2(5)=19.8$, $N=1632$, $p<.01$). Trique listeners' discrimination accuracy was greater at the continuum midpoint than at the continuum boundaries, but the opposite effect was observed for French listeners; discrimination accuracy was greater at the continuum endpoints. For both languages, discrimination accuracy was better for the /2/-/31/ tonal comparison than for the /1/-/31/ comparison.

Fig. 6 shows the discrimination performance for tonal pairs /T2/-/T32/ and /T3/-/T32/. A significant language $\times$ continuum interaction was found for both the /T2/-/T32/ pair ($\chi^2[5,11]=12.7$, $p<.05$) and the /T3/-/T32/ pair ($\chi^2(5)=27.5$, $N=1632$, $p<.001$). Fig. 7 shows discrimination performance between contour tones. A significant language $\times$ continuum interaction was found for both the /T32/-/T31/ tonal pair ($\chi^2(5)=34.0$, $N=1632$, $p<.001$) but not for the /T32/-/T43/ tonal pair ($\chi^2(5)=3.1$, $N=1632$, $p=.68$).

Despite strong differences in the discrimination performance by Trique and French listeners, no general effect of language on *d-prime* (test 3) was observed for any of the tonal comparisons (F(1, 1584)=1.82, $p=.18$). However, a small, but significant language $\times$ tonal comparison interaction was observed (F(7, 1584)=3.0, $p<.05$, with Bonferroni correction). Discrimination accuracy differed across language groups, but these differences did not affect listener sensitivity to a substantial degree.

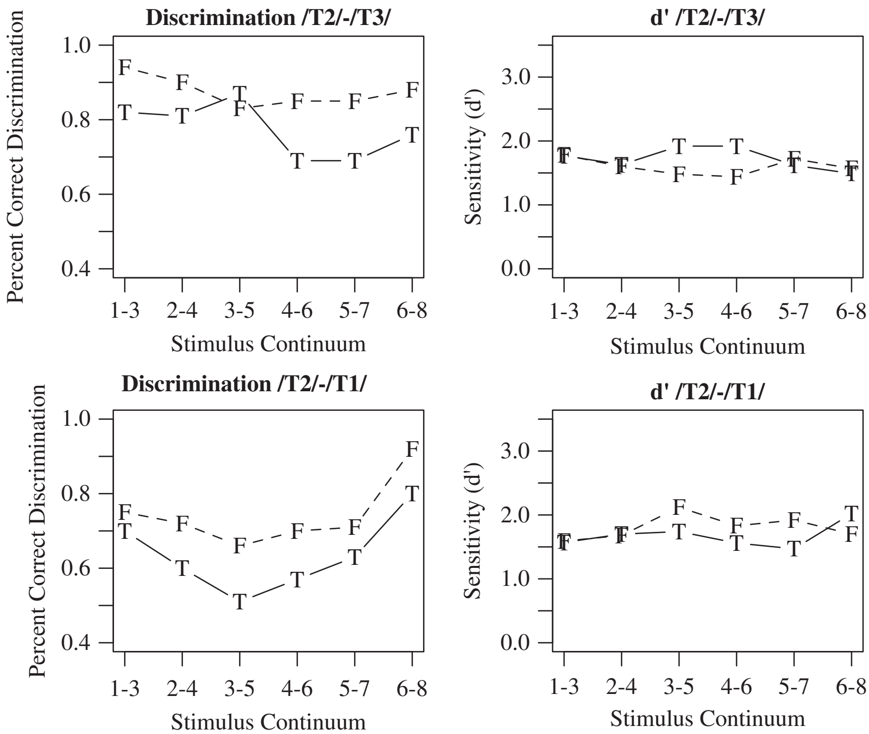

Fig. 4. Discrimination performance, tonal pairs T3-T2 and T2-T1.

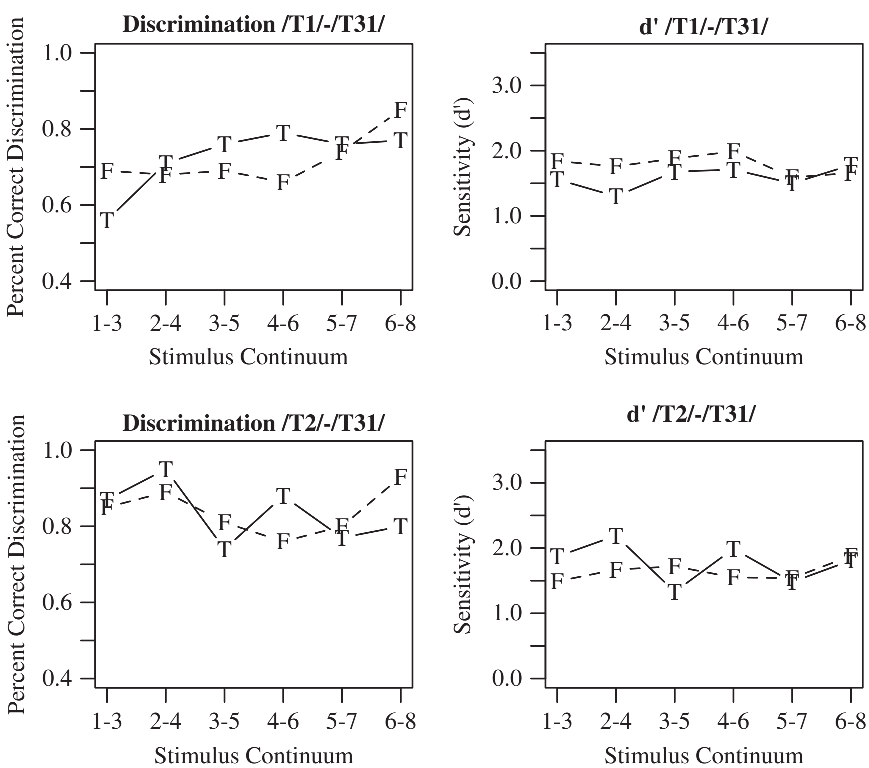

Fig. 5. Discrimination performance, tonal pairs T1-T31 and T2-T31.

2.2.2. Peakedness

With respect to the peakedness of the discrimination function (test 4), a significant main effect of language was found ($F(1, 256)=42.1$, $p<.001$). On average, Trique listeners were 12.5% better at tonal discrimination between categories than within categories. French listeners performed only 2.4% better at tonal discrimination between categories than within categories. A significant main effect of Tonal Comparison was also found ($F(7, 256)=3.8$, $p<.001$). Post-hoc analysis using Tukey's Honest Significant Difference Test showed a significant effect only between tonal comparisons /T2-T1/ vs. /T1-T31/ and /T3-T32/ vs. /T1-T31/ ($p<.01$). Peakedness was lowest for the /T2-T1/ and /T3-T32/ pairs (higher within-category tonal discrimination), while it was highest for the /T1-T31/ pair. In addition, a small,

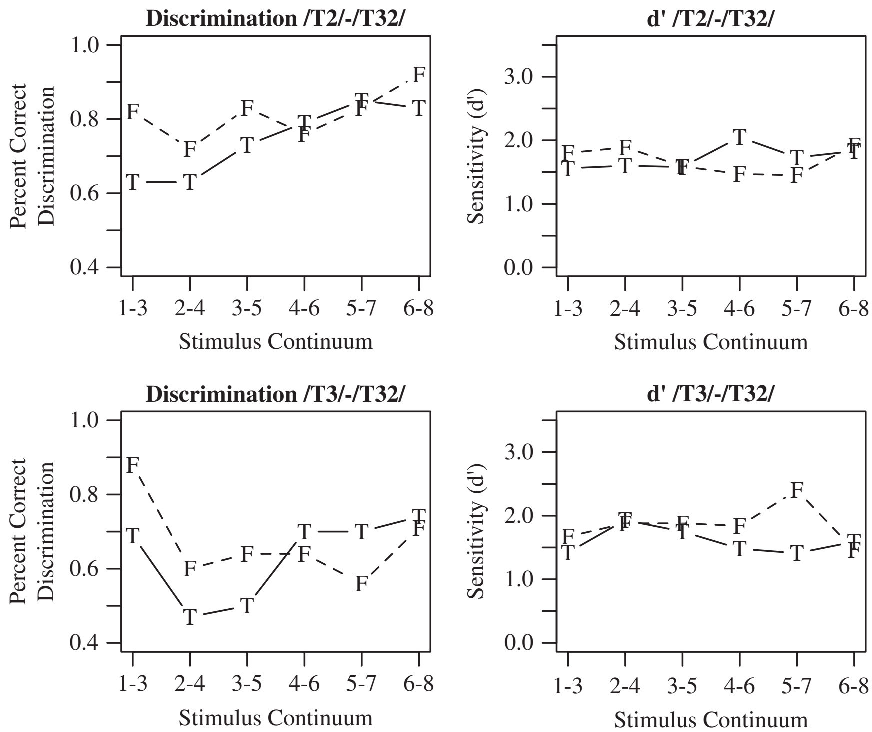

Fig. 6. Discrimination performance, tonal pairs T2-T32 and T3-T32.

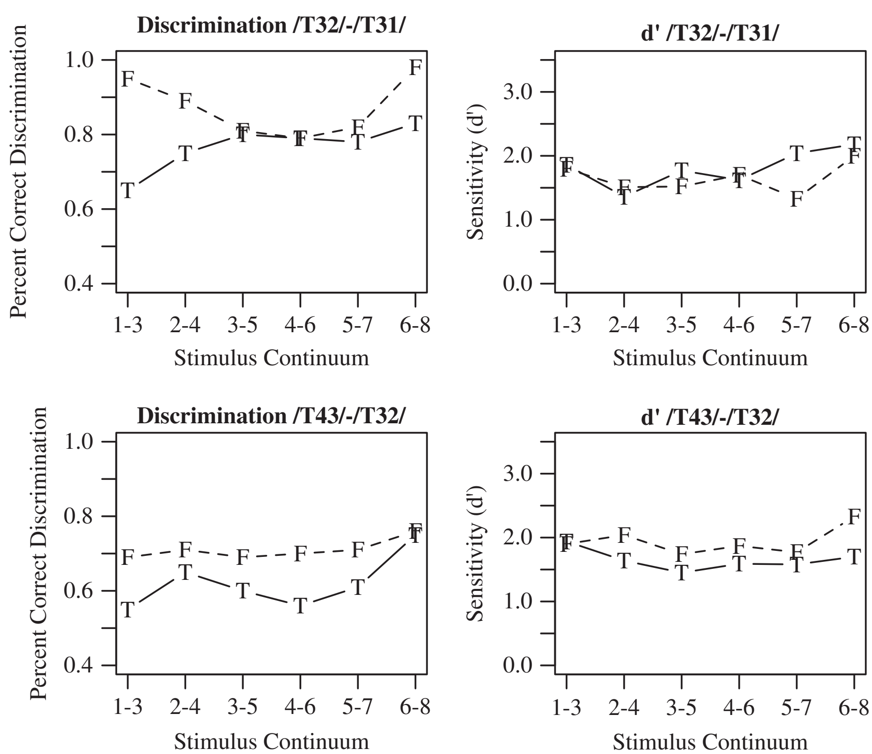

Fig. 7. Discrimination performance, tonal pairs T32-T31 and T43-T32.

but significant, language X tonal comparison interaction was found ($F(7, 256)=2.2$, $p<.05$). Yet, the results from a post-hoc analysis using Tukey's Honest Significant Difference Test found a significant difference only for tonal pair /T1-T31/ ($p<.05$), where the greatest difference in peakedness between language groups occurred.

2.3. Summary and extended analysis

For six of the eight tonal comparisons, Trique listeners showed greater discrimination accuracy between tonal categories than within tonal categories. French listeners were more sensitive to pitch differences within tonal categories than between tonal categories. For Trique listeners, discrimination accuracy was influenced by their knowledge of the linguistic categories associated with each tone. Yet, for all listeners, greater within-category bias in discrimination accuracy may have resulted from the method of resynthesis. Stimuli between the endpoints of each continuum were resynthesized, while endpoint stimuli were not. French listeners showed greater sensitivity to these non-linguistic differences (resynthesized vs. natural stimuli) than Trique listeners did. These boundary effects reflect a sub-phonemic difference between stimuli. Despite such differences, Trique listeners seemed to ignore subphonemic differences, similar to findings in Stagray and Downs (1993).

For two tonal pairs (/T2-T1/, /T43-T32/), no significant peaks in discrimination were observed. In these cases, Trique listeners' discrimination performance followed a similar trajectory to that of French listeners. One explanation for these exceptions may lie in the particular acoustic cues which differentiated the tonal contrasts. Listeners are attentive to a number of acoustic cues in their perception of tonal contrasts, such as average pitch, pitch slope, and pitch direction (Gandour & Harshman, 1978; Liu & Samuel, 2004; Xu, 2004). If we consider these three tonal cues in relation to the phonetic differences between the tones in the discrimination task, a pattern emerges. In Table 3, we observe that the tonal comparisons with fewer contrastive cues did not show evidence of categorical-type peaks in discrimination. Average pitch level here was considered the "same" between tones if their average difference in pitch was less than 1 semitone. Six tonal pairs are distinguished by average pitch. Among these, three pairs are distinguished by this acoustic cue alone: pairs /T3-T2/, /T2-T1/, and /T43-T32/, the latter two for which no significant peaks in discrimination were found across languages. Moreover, since the stimuli were not resynthesized using an acoustic scale, the psychoacoustic distance differed between tonal comparisons. This may also have contributed to differences in discrimination accuracy between tonal comparison blocks.

In order to test these potential confounds, both these cues and psychoacoustic distance were examined in relation to discrimination accuracy. The psychoacoustic distance between stimuli was included in a logistic regression model along with Language and Tonal Comparison as factors. This was evaluated separately against *within*-category and *between*-category discrimination accuracy. Psychoacoustic distance was calculated by taking the average difference in semitones for each stimulus pair in each tonal comparison block. Significant effects of psychoacoustic distance on both *within* ($\chi^2(16) = 103.9, N = 13,056,\ p < .001$) and *between* category discrimination accuracy were found ($(\chi^2)(16) = 61.7$, $N = 13056$, $p < .001$). In general, tonal stimuli with greater psychoacoustic differences were more accurately discriminated by all listeners. Tonal pairs /T2-T1/ and /T3-T32/ were the most psychoacoustically similar, so they were perceived less accurately than the other tonal pairs. A significant language × psychoacoustic distance interaction was observed for both *within* and *between* category discrimination as well, ($\chi^2(8) = 21.7$, $N = 13,056$, $p < .01$; $\chi^2(8) = 21.6$, $N = 13,056$, $p < .01$). As expected, Trique listeners were slightly less sensitive to psychoacoustic differences between stimuli than French listeners were. In general, the worst discriminated tonal pairs were among those which were most psychoacoustically similar. Table 4 compares discrimination performance and psychoacoustic distance for each language group.

Table 3
Contrastive cues among Trique tones.

Tonal comparison	Cues			Cues
	Average pitch	Pitch slope	Pitch direction	
/3/-/2/	Different	Same	Same	1
/2/-/1/	Different	Same	Same	1
/43/-/32/	Different	Same	Same	1
/32/-/31/	Different	Different	Same	2
/3/-/32/	Same	Different	Different	2
/2/-/32/	Different	Different	Different	3
/2/-/31/	Different	Different	Different	3
/1/-/31/	Same	Different	Different	2

Table 4
Psychoacoustic differences between tonal comparisons.

Tonal comparison	Average psychoacoustic distance (semitones)	Discrimination accuracy	
		Trique (%)	French (%)
/32/ - /31/	1.02	76.8	87.2
/2/ - /3/	0.81	77.2	87.5
/43/ - /32/	0.75	62.1	71.1
/2/ - /32/	0.64	74.3	81.1
/2/ - /31/	0.63	83.5	83.4
/1/ - /31/	0.58	72.5	72.0
/2/ - /1/	0.46	63.4	74.2
/3/ - /32/	0.25	63.2	67.2

The effect of the average pitch, pitch slope, and pitch direction on discrimination accuracy were each evaluated separately in a two-way logistic regression model with language treated as a between subjects factor and each of the different cues as within subjects factors. Separate models evaluated *within* category and *between* category discrimination accuracy for each cue. A significant effect of pitch slope on discrimination accuracy was found, but only on *between* category discrimination accuracy ($\chi^2(2) = 36.2$, $N = 13,056$, $p < .001$). This tonal cue is used to distinguish contour from level tones. Listeners were more accurate at tonal discrimination between a contour and a level tone than between two level tones or two contour (falling) tones. A significant interaction of language × slope was also found ($\chi^2(1) = 23.5$, $N = 13,056$, $p < .001$). This finding reflects the observation that only Trique listeners were sensitive to the slope of the pitch contour. Their *between* category discrimination accuracy was 74.0% correct for tones of differing slope values but 67.6% correct for tones with the same slope. French listeners' accuracy was 78.3% correct for tones with differing slope values and 77.6% correct for tones of the same slope. Pitch direction was a significant predictor of discrimination accuracy, but only for the *between* category continua ($\chi^2(2) = 25.8$, $N = 13,056$, $p < .001$). Except for the tonal contrast between /32/ and /31/, this cue largely overlaps with the slope cue.

A significant effect of pitch level was found for both *within* and *between* category discrimination accuracy ($\chi^2(2) = 88.8$, $N = 13,056$, $p < .001$; $\chi^2(2) = 29.5$, $N = 13,056$, $p < .001$). Listeners were more accurate perceiving tonal contrasts with different average pitch levels than with the same pitch level. A significant language × pitch level interaction was found in the *between* category group ($\chi^2(1) = 15.1$, $N = 13,056$, $p < .001$). This finding reflects the observation that French listeners were more sensitive to pitch level as a cue than Trique listeners. Discrimination accuracy among French listeners for tone contrasts with different average pitches was 80.8%, but 69.6% for tone contrasts with the same average pitch level. The difference among Trique listeners was smaller (72.9% and 67.6%). The psychoacoustic distance between stimuli and certain tonal cue differences accounts for some of the variability in discrimination performance among the different tonal comparisons. Yet, these cues are less important in the discrimination of *within* category differences than *between* category differences.

3. Experiment 2: tonal identification with Trique and French listeners

To examine the relationship between the discrimination and identification data, an identification task was carried out. This experiment was similar to the discrimination task, as it also used an AXB (2AFC) design, closely following the methods used in Hallé et al. (2004). The experiment here (and that of Hallé et al.) differed from typical identification tasks in two important ways. First, both Trique and French subjects participated in the experiment. As an identification task examines long-term categorical representations of contrasts, it is uncommon to ask non-native listeners to participate in such a task. However, this method has the advantage of offering a comparison with native speaker identification performance. Second, unlike identification tasks which present a single stimulus per trial which listeners must categorize, the task here included three stimuli per trial which accorded with the structure of an AXB task. The "A" and "B" stimuli served as prototypical categorical comparisons for stimulus "X" for both subject populations.

3.1. Method

3.1.1. Stimuli

The same stimuli created for use in experiment one were used in the identification task. As in the previous experiment, continua were created for six of the seven tonal comparisons. Target words were embedded in a carrier phrase: [ka^{3}tah^3 __] '*Someone said __*'. However, the order and step-size were different in the identification task (see below).

3.1.2. Procedure

In each trial, subjects listened to three stimuli, the first and last of which corresponded to the tonal continuum endpoints and the middle of which corresponded to the test stimulus. Like the discrimination task, each continuum consisted of eight steps. Yet, unlike the previous experiment, the target stimulus consisted of any of the eight possible continua steps. Instructions for the identification experiment were given to Trique subjects in Spanish and to French subjects in French. However, the stimuli were presented without text. As each trial consisted of three stimuli, each stimulus was matched with one of three illustrations. The continuum endpoints were associated with pictures depicting the word's meaning and the test stimulus was associated with a still image of the face of a speaker. For the Trique listeners, subjects were instructed to determine which phrase the speaker's face was repeating. Each stimulus was tested four times (twice for each presentation order: BXA/AXB), for a total of 32 trials per tonal comparison block (8), for a total of 256 stimuli.

All stimuli were presented using a script written for Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993) and presented on the author's Mac iBook G4 laptop. All subjects were seated in a quiet room at a desk facing the author's computer. Each subject listened to the trials over dynamic headphones and responded by pressing one of the two buttons on the keyboard corresponding to the left or right stimulus pair (X="A" *left* and X="B" *right*). The test phase was preceded by a practice block of 32 trials, which consisted of four trials from each of the eight tonal continua. Following this practice block, listeners were asked if they understood the instructions of the experiment. Those who did were told to proceed in the test trials.

3.1.3. Subjects

Seventeen Trique subjects participated in the experiment. One subject was eliminated due to not having completed the experiment. All were native speakers of the San Martín Itunyoso dialect of Trique. For 13 of the 17 Trique subjects, the experiment took place in the town of San Martín Itunyoso. For the remaining subjects, the experiment took place in a quiet room in the town of Tlaxiaco, Mexico. Thirteen French subjects participated in the identification task. For the French subjects, the experiment took place in a laboratory at Université Lyon 2, in Lyon, France. All subjects were between 18 and 45 years old at the time of the experiment. No subject reported any history of speech or hearing disorders. Only two of the Trique subjects who participated in the discrimination task also participated in the identification task. None of the French subjects who participated in the identification task participated in the discrimination task. This differs from typical categorical perception studies where all subjects participate in discrimination and identification tasks. However, given the field conditions of this study, it was not possible to recruit Trique subjects for both experiments.

3.2. Results

For each tonal pair, identification performance was statistically analyzed using a two factor logistic regression model with language and continuum step as within subjects factors. A χ^2 test was run on each logistic model to determine goodness of fit. While each tonal pair was statistically analyzed individually (below), a general effect of language on identification performance was found for the pooled tonal pairs ($\chi^2(64) = 136.1$, $N = 7168$, $p < .001$), but only a small interaction of continuum step X language was observed ($\chi^2(56) = 80.5$, $N = 7168$, $p < .05$).

A significant effect of language on identification performance was found for five of the eight tonal pairs. Figs. 8 and 9 show the identification task performance for Trique and French listeners. In Fig. 8, an effect of language was only found for tonal pair /T2-T1/ ($\chi^2(8) = 18.2$, $N = 896$, $p < .05$). Yet, this effect did not reflect any sensitivity to tonal categories. Rather, Trique subjects were significantly less likely to categorize stimuli as tone /2/ as French subjects were. For tonal pairs /T3-T2/, /T1-T31/, and /T2-T31/, no significant language effects were observed. Trique subjects categorized stimuli on these continua similarly to French subjects.

In Fig. 9, the effect of language on identification was strongest for the /T3-T32/ pair, $\chi^2(8) = 37.3$, $N = 896$, $p < .001$, where a steeper slope was found in the identification function for Trique listeners than for French listeners. A significant main effect of language was also found for the /T2-T32/tonal pair ($\chi^2(8) = 18.0$, $N = 896$, $p < .05$), for the /T32-T31/tonal pair ($\chi^2(8) = 16.1$, $N = 896$, $p < .05$), and for the /T43-T32/ tonal pair ($\chi^2(8) = 19.5$, $N = 896$, $p < .05$).

The location of the identification cross-over for each language was determined by interpolating the location of category change for each individual speaker for each tonal comparison. These values were then statistically analyzed with a two-factor ANOVA with language and tonal comparison as factors. A small, but significant main effect of language, was observed ($F(1, 178) = 4.75$, $p < .05$). The identification cross-over occurred later on the tonal continuum for Trique listeners than for French listeners. A significant main effect of tonal continuum was also found ($F(7, 178) = 5.36$, $p < .001$). The change in identification occurred earlier for certain tonal comparisons, like /T2-T3/ than for others. Table 5 shows the identification cross-over locations for each of the tonal comparisons.

3.3. Relationship between discrimination and identification

The degree of categorical perception among listeners was determined by comparing predicted discrimination accuracy from identification performance to the observed discrimination scores

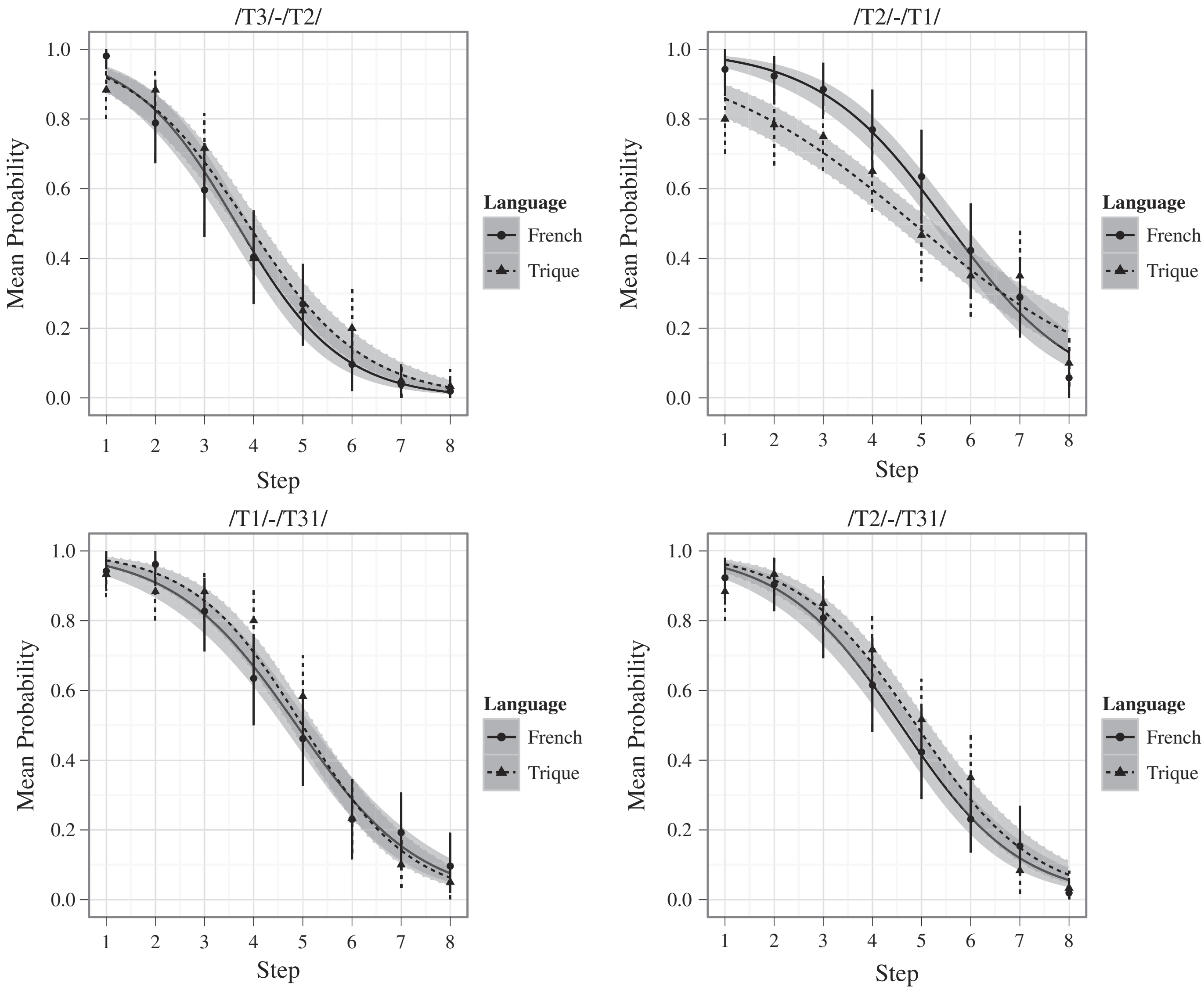

Fig. 8. Identification performance for Trique and French listeners, first four continua. The width of the vertical bar represents a confidence interval with an alpha level set at .05. The gray outline to each curve represents the standard error of the logistic fit.

(Pisoni, 1973; Xu et al., 2006). Predicted discrimination scores were computed using Eq. (1) (Pollack & Pisoni, 1971). Here, P_A represents the proportion of stimuli identified as "A" at continuum step i and P_B represents the proportion of stimuli identified as "A" at continuum step $i+2$

$$P^* = [1+(P_A - P_B)^2]/2 \quad (1)$$

In most studies of categorical perception, the discrimination performance of individual listeners is compared to their identification performance. In the work presented here, the subject pools differed between the discrimination and identification tasks. As a result, a direct comparison of the relationship between discrimination and identification was not possible. Instead, the discrimination data were converted to a matrix of probabilities. These were compared to the predicted discrimination accuracy in a series of matched t-tests. If significant, the tests indicate that observed discrimination performance is better than predicted from identification. If not significant, these tests indicate that observed discrimination performance is not better than predicted.

Statistical analyses were divided into *between*-category and *within*-category discrimination. For *between*-category differences, the results found that observed discrimination accuracy was greater than predicted accuracy across both languages ($t(1921) = -15.6$, $p < .001$). Discrimination accuracy was much higher than predicted from identification performance (73.1% vs. 61.1%). The same pattern was observed for *within*-category differences ($t(779) = -25.7$, $p < .001$). Observed *within*-category discrimination accuracy was much higher than predicted (79.9% vs. 55.2%). The predicted and observed discrimination performance for each language and tonal comparison are given in Figs. 10 and 11.

The differences between predicted and observed discrimination accuracy were compared for each language and each tonal comparison. For French listeners, *between*-category predicted accuracy was significantly lower than observed accuracy across all tonal comparisons ($t(1050) = -14.6$, $p < .001$), (60.2% vs. 74.8%). The same pattern was observed for Trique listeners ($t(824) = -7.5$, $p < .001$), (61.9% vs. 70.7%). *Within*-category predicted accuracy was also significantly lower than observed accuracy for both language groups: French ($t(490) = 24.5$, $p < .001$) and Trique ($t(300) = 12.1$, $p < .000$). However, French listeners were more sensitive to *within*-category differences than Trique listeners were (73.4% vs. 84.5%).

The location of predicted peak locations was estimated for all speakers who participated in the identification task. This was

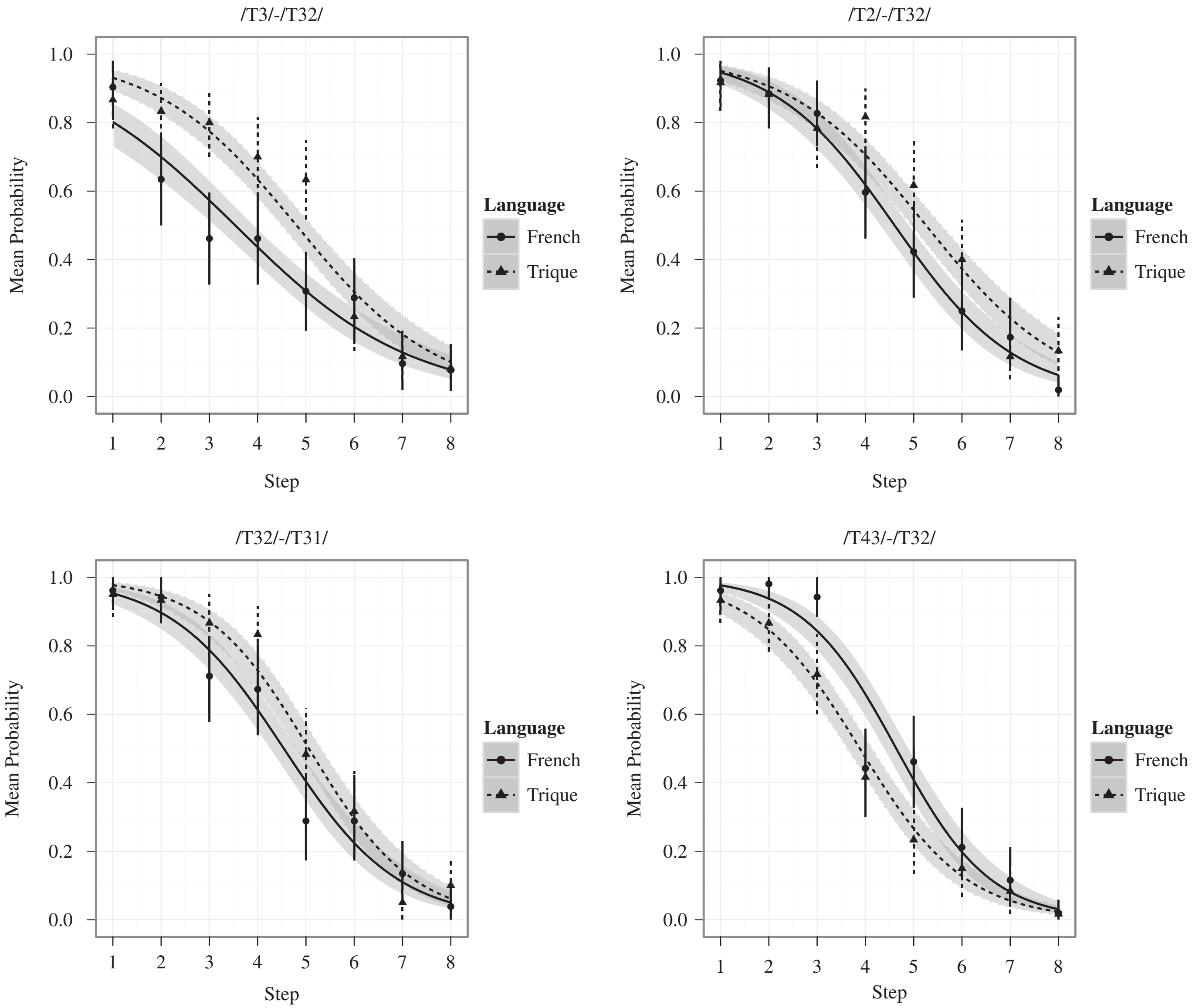

Fig. 9. Identification performance for Trique and French listeners, second four continua. The width of the vertical bar represents a confidence interval with an alpha level set at .05. The gray outline to each curve represents the standard error of the logistic fit.

Table 5
Location of identification cross-over between languages.

Tonal continuum	Identification cross-over (eight point scale)	
	Trique	French
T2-T3	3.90	3.78
T2-T1	4.90	5.02
T1-T31	5.04	4.61
T2-T31	5.11	4.63
T2-T32	5.37	4.77
T3-T32	4.92	4.33
T32-T31	5.07	4.58
T43-T32	4.17	4.65

compared to the observed peak location from speakers who participated in the discrimination task. The difference between the predicted and observed peak locations by language are shown in Table 6. For five of the eight tonal continua, the predicted and the observed peak locations matched more closely (lower difference score) for Trique listeners than for French listeners. For three of the continua, T2-T31, T3-T32, and T43-T32, the opposite pattern was observed. In general, the predicted and observed peak locations were slightly closer for Trique listeners than for French listeners.

There was a significant difference between predicted and observed discrimination accuracy for every tonal comparison for French listeners. Although, for Trique listeners, the difference between predicted and observed accuracy varied by tonal comparison. For tonal pairs /2/-/1/ and /43/-/32/, no differences in predicted and observed *between*-category discrimination accuracy were found. This is notable, as no language $\times$ continuum effect was observed for these tonal pairs in discrimination (see Section 2.2.1). If these tones were perceived more psychoacoustically, without any apparent sensitivity to categorical boundaries, then the predicted and observed discrimination accuracy would more closely correspond. In addition, no differences in predicted and observed *between*-category discrimination accuracy were observed for tonal pairs /3/-/32/ and /2/-/32/. This finding may more closely reflect categorical perception between these tones. However, tones /3/ and /32/ were the most psychoacoustically similar pair and *between*-category discrimination accuracy was quite low for Trique listeners. In general, discrimination accuracy

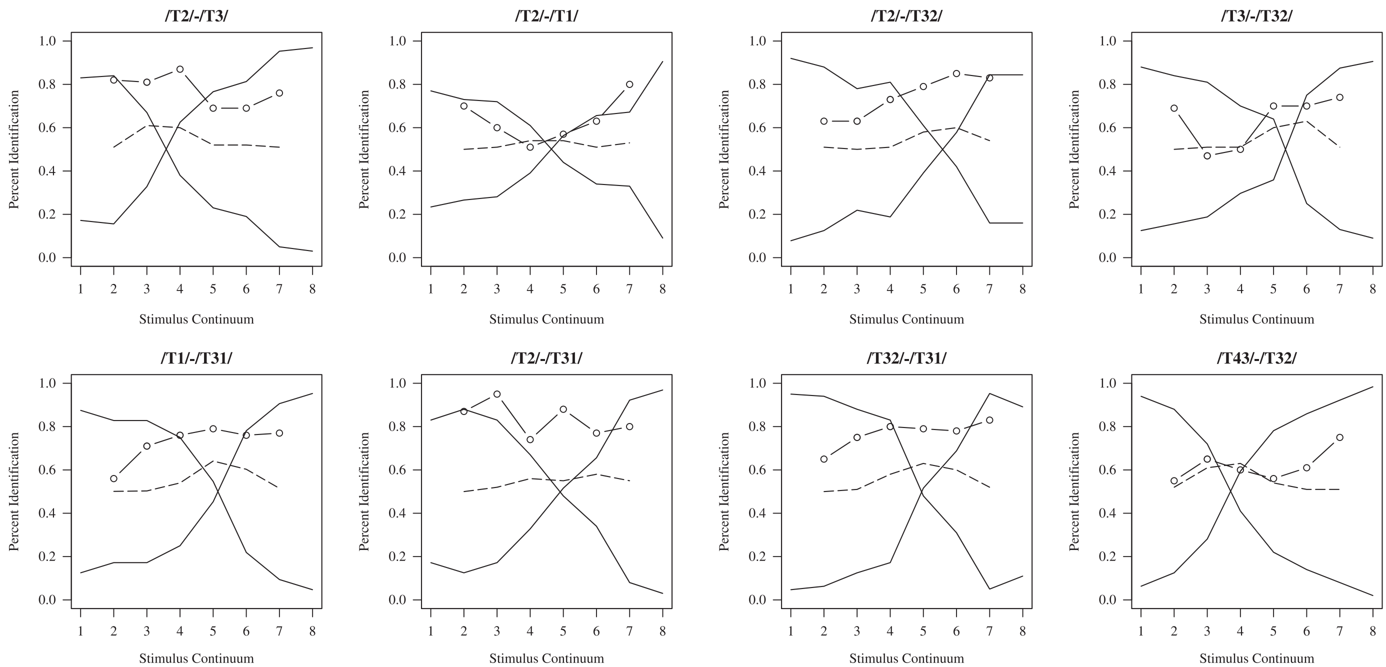

Fig. 10. Observed discrimination performance (circles) and predicted discrimination performance (dashes) for Trique listeners. Data are plotted on the observed identification performance.

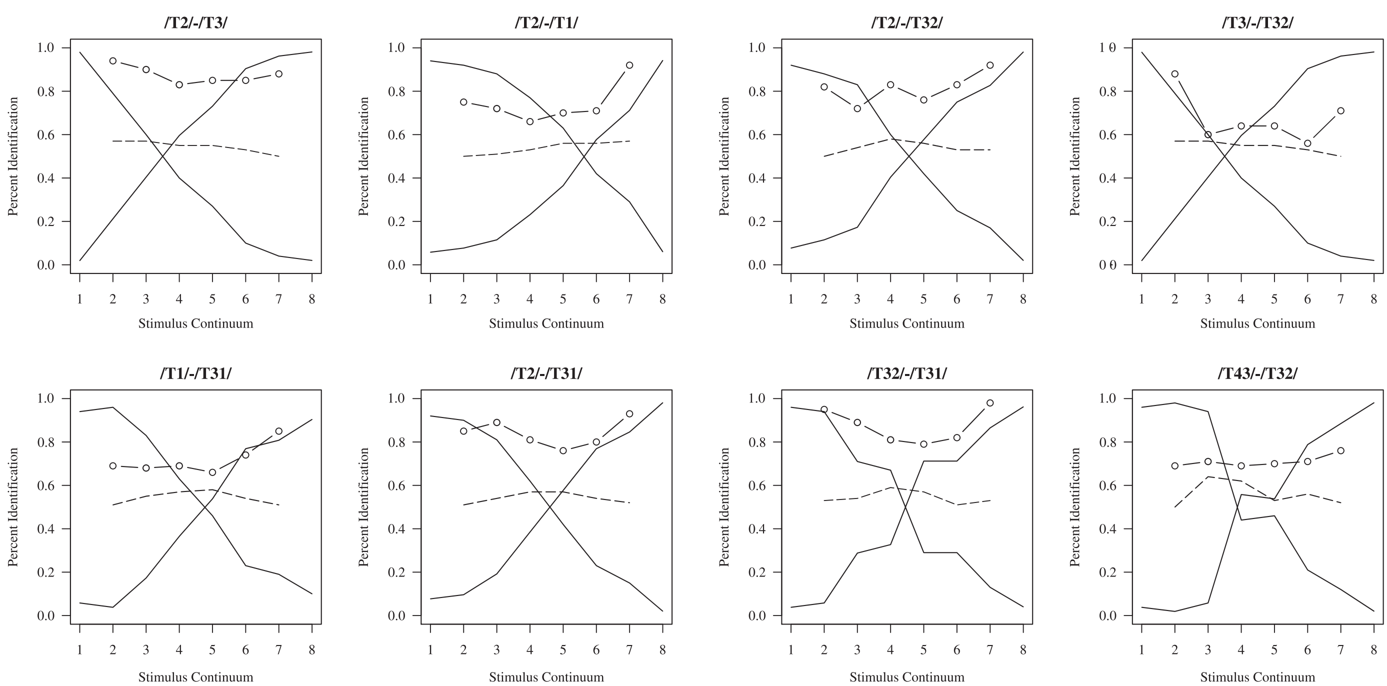

Fig. 11. Observed discrimination performance (circles) and predicted discrimination performance (dashes) for French listeners. Data are plotted on the observed identification performance.

predicted from the identification experiment did not correspond to the observed accuracy from the discrimination task. Discrimination accuracy was underestimated.

4. Discussion

4.1. Sensitivity to tonal categories in tone perception

In relation to the criteria for categorical perception stated in Section 1.1, the findings from these experiments are mixed. First, Trique listeners did not show more abrupt identification for tonal contrasts than French listeners. The differences in the labeling probabilities between languages resulted from differences in the categorical boundary, not in the slope of the identification function. Second, while the peakedness of the discrimination function varied between the language groups, observed peaks in discriminability did not correspond to identification boundaries. Third, within-category discrimination accuracy was much lower for Trique listeners than between-category discrimination accuracy. The opposite pattern held for French listeners. For all listeners, discrimination remained above chance level within the tonal

Table 6
Difference between predicted and observed peaks in discrimination.

Tonal continuum	Differences in peak location (in steps)	
	Trique	French
T2-T3	.051	.104
T2-T1	.291	.415
T1-T31	.194	1.142
T2-T31	1.161	.547
T2-T32	.174	.624
T3-T32	.380	.064
T32-T31	.249	.453
T43-T32	.807	.439

categories. Finally, little correspondence was found between the predicted and observed peaks in discrimination. At first glance, the results from the discrimination experiment appear similar to those found for the categorical perception of tone. Yet, the identification data do not suggest a categorical mode of perception. Instead, a strong language bias occurred in discrimination and identification, but the behavior of the former was not mediated by performance in the latter.

Typically, in studies which have found no categorical perception of tone, no differences in between-category discrimination accuracy are observed (Abramson, 1979; Avelino, 2003; Connell, 2000). However, the present study differs in important ways from this previous work. In each of these studies where no categorical effect was observed, researchers investigated level tones in citation form. In the current study, stimuli were placed within a natural speech context. This context offers an extrinsic comparison for native listeners. Such comparisons are necessary for the categorization of steady-state phenomena, like level tones and steady state vowels (Mirman, Holt, & McClelland, 2004). Steady-state phenomena are less likely to be categorically perceived when presented in isolation. Furthermore, since contextual variability is much larger for vowels (and tones) than for consonants, extrinsic effects are much stronger. It is more likely that listeners will rely on these contextual cues for category identification. If none are given, there is no categorical perception (Xu et al., 2006).

There were substantial psychoacoustic effects on discrimination performance. Tonal contrasts differing in only one acoustic cue were less accurately discriminated than those differing in multiple cues. For certain tonal pairs differing only in pitch level (/T2-T1/ and /T43-T32/), no language effect on tonal discrimination was observed. One explanation for this effect is that minimally-cued tonal contrasts like these may require more informative extrinsic contexts to produce an observable language bias. The carrier sentence into which these comparisons were embedded was phonologically informative, as the stimuli were adjacent to words with /T3/. However, the context did not provide any relevant semantic information by which native listeners could distinguish these pairs.

There is also a strong possibility that native Trique listeners are simply less sensitive to these tonal contrasts than to other contrasts. Like tonal pair /T2-T1/, the /T3-T2/ contrast only differs in pitch level, but it has a much larger functional load. For instance, tones /1/ and /2/ are rarer in the Trique lexicon than tone /3/. In a corpus of 1831 words, tone /1/ occurs in only 3.2% of words and tone /2/ occurs in only 3.9% of words. In contrast, tone /3/ occurs in 18.1% of words. Moreover, the difference between perfective and potential aspect on Trique verbs is minimally marked by a change from /T3/ to /T2/ (DiCanio, 2010a). Listeners may be more sensitive to the /T3-T2/ contrast than to the /T2-T1/ contrast because the former is morphologically contrastive. The /T43-T32/ contrast is also functionally different than other tonal contrasts. Most words with tone /43/ are loanwords from Spanish, many of which do not contrast with phonologically similar words in Trique. Categorical effects in discrimination may be sensitive to the function of the tonal contrast in the language.

The differences in sensitivity to certain tonal cues between Trique and French listeners closely matches previous work on tonal cue sensitivity. Gandour and Harshman (1978) found that Thai and Yoruba listeners were more sensitive to F_0 slope as a tonal cue than English listeners. In the study here, Trique listeners were more accurate than French listeners in discriminating between tonal pairs which differed in F_0 slope, while French listeners were more accurate than Trique listeners in discriminating between tonal pairs which differed in average F_0. This finding suggests that greater perceptual sensitivity to F_0 slope may be a characteristic of native tone language listeners.

In general, tonal stimuli that were more psychoacoustically similar were discriminated less accurately. This was true for both Trique and French listeners. This effect may have resulted from using an acoustic scale for pitch resynthesis instead of a psychoacoustic scale using ERB (equivalent rectangular bandwidth), the method used in similar studies of tone perception (Hallé et al., 2004; Xu et al., 2006).[3] For instance, discrimination may have been more difficult in tonal pairs /T2-T1/ and /T3-T32/, as the frequency difference between steps in each of these continua was only 3.7 Hz, or 0.45 semitones. This difference approaches the just noticeable difference (JND) observed for pitch contrasts, which is typically 2–3 Hz (Hillenbrand & Houde, 1996; Vaerenberg et al., 2011).

The significance of these factors in tonal discrimination and the lack of abrupt category boundaries in identification suggest that Trique listeners rely on the raw acoustic differences in tonal categorization but are also more sensitive to non-prototypical tonal exemplars straddling the categorical boundaries. French listeners relied more on absolute pitch differences in their discrimination of Trique tones. This finding runs counter to the claim that native tone language listeners rely more on absolute pitch differences in tonal discrimination due to their experience with linguistic tone (Peng et al., 2010).

The results from the identification task suggest that there are language differences in the location of the perceptual boundaries for tone. However, differences in the slope of the identification function were found for only half the tonal comparisons. This finding differs slightly from findings in the literature stating that tonal identification curves are always steeper for native speakers of tone languages than for non-native speakers (Hallé et al., 2004; Peng et al., 2010; Xu et al., 2006). The steepness of the identification curve is generally considered a good measure of categorical perception, but it was not a strong indicator in the present study.

Sensitivity to tonal category boundaries in discrimination accuracy was observed for within-category stimuli for Trique listeners. In typical discrimination tasks, stimuli are presented in isolation with a short ISI (100–200 ms). The purpose of a short interval is to ensure that listeners will have access to auditory memory. Pitch information decays much more slowly than the transitional cues used for consonants, so it is particularly useful in the perceptual recall of tones. In the current study, the ISI was set to 500 ms. Moreover, each target stimulus was preceded by a carrier word of 500 ms. The long duration between stimuli (1000 ms) in the discrimination experiment would predict a greater reliance on phonetic memory in tonal discrimination. This is precisely what was observed. Native Trique listeners were less accurate at discriminating within-category phonetic

[3] Note that the large number of tonal contrasts in Trique differ in psychoacoustic distance from each other. If an ERB scale were used, the number of steps would have had to differ for each tonal continuum. An acoustic scale obviates this.

differences than between-category differences because they relied on more categorical memory representations.

Peaks in discrimination sensitivity for Trique listeners were shallower than those typically observed in discrimination studies examining consonantal contrasts, cf. Harnad (1987). A similar result has been observed for Taiwanese Mandarin listeners (Hallé et al., 2004). The explanation for this phenomenon lies in how certain types of processing memory are used by listeners in speech discrimination. *Auditory memory* has a short time-span (between 200 and 300 ms) and does not involve access to categorical long-term memory representations of speech. *Phonetic memory* lasts longer and involves a categorization of the stimulus based on long-term memory representations (Gerrits & Schouten, 2004; Pisoni, 1973; Repp, 1984). Tonal targets span entire syllables with their most significant acoustic properties occurring on the vowel. Consonants have more abrupt acoustic properties which span a short temporal interval. The decay of changing acoustic information occurs so quickly for consonants that listeners cannot rely on auditory memory, but must rely on phonetic memory (Gerrits & Schouten, 2004). Abrupt and canonically categorical peaks in discrimination occur for contrasts which span a short temporal interval, but broader, less canonically categorical peaks in discrimination occur for contrasts spanning longer temporal intervals. Long interval contrasts like tone rely more on an auditory mode of processing than short interval consonantal contrasts. As a result, one predicts shallower peaks to occur for tonal discrimination than for consonantal discrimination.

4.2. *Tone perception among native and non-native listeners*

Across all tonal comparisons, French listeners showed greater discrimination accuracy than Trique listeners. This finding runs counter to claims in Burnham et al. (1996) and Hallé et al. (2004), where native tone language listeners are argued to perform better in pitch discrimination tasks. However, it accords with findings in Stagray and Downs (1993), where the authors found that Mandarin listeners were *less* sensitive to within-category pitch differences than English listeners were. The Trique listeners' data resembles the native Mandarin listeners data in this respect.

Trique listeners may have also had lower discrimination performance than French listeners due to cultural differences between the subject populations. While most of the French subjects were college-age students who had some familiarity with the expectations of subjects in experimental research, the Trique subjects were more varied. They had a larger age-range (18–45 years old) than the French subjects (18–34 years old). Many Trique subjects had received college or secondary school education, but certain subjects had not. Furthermore, there is little or no familiarity with research in higher education among Trique speakers. Finally, the discrimination task was given in Spanish as there is no native-language literacy among the Trique listeners. Unlike the identification task, which used pictures instead of written words, the discrimination task required subjects to respond by clicking on words written in Spanish (*izquierda* 'left' and *derecha* 'right'). Trique listeners may have been biased to provide Spanish-like responses since the instructions of the discrimination task were given in Spanish. However, there is reason to question this. While it is possible that the Trique listeners paid less attention to the pitch differences they heard (since Spanish is non-tonal), the target words did appear embedded in a Trique sentence. Trique speakers frequently switch between Trique and Spanish when they are engaged in conversations with outsiders, speaking Trique between themselves, and responding to the outsider in Spanish. The inclusion of these two languages, side-by-side, is not out of the ordinary in their language experience. Nevertheless, the language of the experiment was not explicitly tested as a factor in these experiments and it cannot be discounted as a possible influence on the ability of Trique listeners to discriminate the tonal contrasts.

Despite these possible cultural differences, the present findings argue that native language listeners do not have an advantage discriminating native contrasts over non-native listeners. Among segmental contrasts, this is well-established (Best & Tyler, 2007). For instance, listeners of American English are better at discriminating certain click consonants than native listeners of click languages (Best, Traill, Carter, Harrison, & Faber, 2003). One explanation given for this effect is that non-native listeners will evaluate stimuli based on its similarity to their native language contrasts (the Perceptual Assimilation Model, PAM) (Best & Tyler, 2007). However, such an explanation less clearly applies here to the perception of tone by French listeners. French has no tonal contrasts and French listeners perform poorly in the perception of accentual contrasts (Dupoux, Pallier, Sebastián-Gallés, & Mehler, 1997, 2008). Instead, the findings here suggest that French listeners use a more purely auditory mode of perception (Frankish, 2008) and perceive tone more psychoacoustically.

5. Conclusions

The series of experiments presented here show evidence for a sensitivity to categorical boundaries in discrimination among Trique listeners and a language bias in identification. In an AXB task examining the discrimination of Trique tonal contrasts with Trique and French listeners, French listeners showed greater discrimination accuracy than Trique listeners. Trique listeners showed significant between-category peakedness in discrimination accuracy and a marked decrease in accuracy in within-category discrimination, in accord with findings in Stagray and Downs (1993). No such abrupt changes in discriminability were observed for French listeners. These findings suggest that Trique listeners are sensitive to linguistic categories in discriminating tonal contrasts. However, no clear relationship between discrimination and identification performance was found. These findings are in agreement with recent work showing categorical effects in tone discrimination but little relationship between discrimination and identification (Francis et al., 2003).

Listeners were sensitive to psychoacoustic differences in the data. The psychoacoustic distance between tonal stimuli and the number of cues which distinguished different tones contributed to their discriminability. Like vowels, one predicts greater psychoacoustic and contextual sensitivity to tonal contrasts (Xu et al., 2006). So, these results are, at least in part, expected. However, the exact role of these extrinsic cues in the categorization of tone is a largely unanswered one. For instance, it is generally understood that the listeners are more sensitive to dynamic cues than static ones (Kewley-Port, Pisoni, & Studdert-Kennedy, 1983; Mirman et al., 2004). The current paper provides some evidence for this hypothesis, as listeners were better discriminating tonal contrasts with a number of cues than contrasts with fewer cues. Though, it remains to be tested in a more controlled manner. The hypothesis predicts that listeners would more poorly discriminate tones differing in only one static dimension (pitch level, intensity) than tones differing in dynamic cues (pitch slope or spectral differences). The answer to this particular question may shed light on how psychoacoustic sensitivities integrate with categorical responses in tone perception, but it remains a question for future research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at http://dx.doi.org.10.1016/j.wocn.2012.05.003.

References

Abramson, A. S. (1962). The vowels and Tones of Standard Thai: Acoustical measurements and experiments. Indiana University Research Center in Anthropology, Folklore, and Linguistics: Bloomington, Publication 20.

Abramson, A. S. (1972). Tonal experiments with whispered Thai. In: A. Valdman, & P. Delattre (Eds.), *Papers in linguistics and phonetics to the memory of Pierre Delattre* (pp. 31–44). Mouton: The Hague.

Abramson, A. S. (1975). The tones of Central Thai: Some perceptual experiments. In: J. G. Harris, & J. Chamberlain (Eds.), *Studies in Tai linguistics* (pp. 1–16). Bangkok: Central Institute of English Language.

Abramson, A. S. (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.

Avelino, H. (2003). Categorical perception of phonemic tone in Yalálag Zapotec. In *Proceedings of the 15th International Congress of the Phonetic Sciences*.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*, 305–330.

Best, C. T., Traill, A., Carter, A., Harrison, K. D., & Faber, A. (2003). !Xóõ click perception by English, Isizulu, and Sesotho listeners. In: M.-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of the phonetic sciences* (pp. 853–856). Barcelona: Causal Productions.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In: M. J. Munro, & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer [computer program] 〈www.praat.org〉.

Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, *37*(1), 79–96.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tone across languages: Evidence for alinguistic mode of processing. In *Proceedings of the 4th international conference on spoken language processing* (Vol. 4, pp. 2514–2517).

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, *62*, 961–970.

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.

Cho, T., & McQueen, J. M. (2006). Phonological versus phonetic cues in native and non-native listening: Korean and Dutch listeners' perception of Dutch and English consonants. *Journal of the Acoustical Society of America*, *119*(5), 3085–3096.

Chuang, C.-K., Hiki, S., Sone, T., & Nimura, T. (1972). The acoustical features and perceptual cues of the four tones in standard colloquial Chinese. In *Proceedings of the 7th international congress of acoustics* (Vol. 3, pp. 297–300). Budapest: Akadémial Kiado.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*(2), 257–271 [Version B57 for OSX, 2010, available at http://psy.cns.sissa.it/http://psy.cns.sissa.it/].

Connell, B. (2000). The perception of lexical tone in Mambila. *Language and Speech*, *43*(2), 163–182.

Cutting, J. E., & Rosner, B. S. (1976). Discrimination functions predicted from categories in speech and music. *Perception and Psychophysics*, *20*(1), 87–88.

DiCanio, C. (2010a). The structure of the Itunyoso Trique Word. Manuscript.

DiCanio, C. T. (2008). *The Phonetics and Phonology of San Martín Itunyoso Trique*. Ph.D. thesis, University of California, Berkeley.

DiCanio, C. T. (2010b). Illustrations of the IPA: San Martín Itunyoso Trique. *Journal of the International Phonetic Association*, *40*(2), 227–238.

DiCanio, C. T. (2012a). Coarticulation between Tone and Glottal Consonants in Itunyoso Trique. *Journal of Phonetics*, *40*, 162–176.

DiCanio, C. T. (2012b). The phonetics of Fortis and Lenis Consonants in Itunyoso Trique. *International Journal of American Linguistics*, *78*(2), 239–272.

DiCanio, C. T., & Cruz Martínez, B. (2010). Chungwì Snáhánj nìh, El Mundo Triqui: Palabras de San Martín Itunyoso. Instituto Nacional de Lenguas Indígenas.

Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A destressing "deafness" in French?. *Journal of Memory and Language*, *36*, 406–421.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, *106*, 682–706.

Francis, A. L., Ciocca, V., King Yu Wong, N., Ho Yin Leung, W., & Cheuk YanChu, P. (2006). Extrinsic context affects perceptual normalization of lexical tone. *Journal of the Acoustical Society of America*, *119*(3), 1712–1726.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*, 268–294.

Francis, A. L., Ciocca, V. C., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception and Psychophysics*, *65*, 1029–1044.

Frankish, C. (2008). Precategorical acoustic storage and the perception of speech. *Journal of Memory and Language*, *58*, 815–836.

Frazier, M. (2009). *The production and perception of Pitch and Glottalization in Yucatec Maya*. Ph.D. thesis, University of North Carolina, Chapel Hill.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*(4), 171–189.

Gandour, J., & Harshman, R. (1978). Cross-language differences in tone perception. *Language and Speech*, *21*(1), 1–33.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics*, *66*(3), 363–376.

Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, *27*, 281–306.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421.

Harnad, S. R. (1987). *Categorical perception: The groundwork of cognition*. Cambridge University Press.

Hillenbrand, J. M., & Houde, R. A. (1996). Role of F_0 and amplitude in the perception of intervocalic glottal stops. *Journal of Speech and Hearing Research*, *39*, 1182–1190.

Hollenbach, B. E. (1984). *The phonology and morphology of tone and laryngeals in Copala Trique*. Ph.D. thesis, University of Arizona.

Hombert, J. M. (1975). Perception of contour tones: an experimental investigation. In *Proceedings of the first annual meeting of the berkeley linguistics society* (Vol. 1, pp. 221–232).

Hombert, J. M. (1976). Perception of tones of bisyllabic nouns in Yoruba. *Studies in African Linguistics* (Supplement 6), 109–121.

Howie, J. M. (1972). Some experiments on the perception of Mandarin tone. In: A. Rigault, & R. Charbonneau (Eds.), *Proceedings of the seventh international congress of phonetic sciences* (pp. 900–904). The Hague: Mouton.

Huang, T. (2004). *Language-specificity in auditory perception of Chinese tones*. Ph.D. thesis, The Ohio State University.

Huang, T., & Johnson, K. (2010). Language-specificity in speech perception: Perception of mandarin tones by native and nonnative listeners. *Phonetica*, *67*, 243–267.

Kewley-Port, D., Pisoni, D. B., & Studdert-Kennedy, M. (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, *73*(5), 1779–1783.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*, 69–72.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception and Psychophysics*, *32*, 542–550.

Lavelle, C. R. (1974). An experimental study of Yoruba tone. *Studies in African Linguistics, Supplement*, *5*, 185–194.

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in cantonese and mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 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.

Lin, H.-B., & Repp, B. H. (1989). Cues to the perception of Taiwanese tones. *Language and Speech*, *32*(1), 25–44.

Liu, S., & Samuel, A. (2004). Perception of mandarin lexical tones when f0 information is neutralized. *Language and Speech*, *47*(2), 109–138.

Macmillan, N. A. (1987). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In: S. R. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 53–85). Cambridge University Press (Chapter 2).

Maddieson, I. (2010). Tone. In: M. Haspelmath, M. Dryer, S. Matthew, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library (Chapter 13, Accessed on 10/27/2010).

Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., & Dooling, R. J. (1976). Discrimination and labeling of noise-buzz sequences of categorical perception. *Journal of the Acoustical Society of America*, *60*, 410–417.

Mirman, D., Holt, L. L., & McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America*, *116*, 1198–1207.

Pastore, R. E. (1987). Categorical perception: Some psychophysical models. In: S. R. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 29–51). Cambridge University Press (Chapter 1).

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*, 616–624.

Peng, S.-h. (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics*, *25*, 371–400.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, *13*(2), 253–260.

Pollack, I., & Pisoni, D. B. (1971). On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, *24*, 299–300.

R Development Core Team, Vienna, A. (2009). R: A language and environment for statistical computing [computer program]. R Foundation for Statistical Computing 〈http://www.R-project.org〉.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In: N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice*, vol. 10 (pp. 243–335). New York: Academic Press.

Rosen, S., Howell, P., 1987. Auditory, articulatory, and learning explanations of categorical perception in speech. In *Categorical perception: The groundwork of cognition* (pp. 113–160, Chapter 3). Cambridge University Press.

So, C. K. (2006). Perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. In: P. Warren, & C. I. Watson (Eds.), *Proceedings of the 11th Australian international conference on speech science & technology*. New Zealand: University of Auckland.

Stagray, J., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and nontone language. *Journal of Chinese Linguistics*, *21*, 143–163.

Vaerenberg, B., Pascu, A., Del Bo, L., Schauwers, K., De Ceulaer, G., Daemers, K., Coene, M., & Govaerts, P. J. (2011). Clinical assessment of pitch perception. *Otology and Neurotology*, *32*(5), 736–741.

Whalen, D. H., & Xu, Y. (1992). Information for mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*, 25–47.

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in cantonese level tones. *Journal of Speech and Hearing Research*, *46*, 413–421.

Xu, C., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, *33*(2), 165–181.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, *96*(4), 2240–2253.

Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics*, *5*, 757–797.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, *120*(2), 1063–1074.

Zsiga, E., & Nitisaroj, R. (2007). Tone features, tone perception, and peak alignment in Thai. *Language and Speech*, *50*, 343–383.