Auditory-Visual Speech Processing
(AVSP) 2011
Volterra, Italy
September 1-2, 2011

Auditory-Visual Discrimination and Identification of Lexical Tone Within and Across Tone Languages

Denis Burnham, Virginie Attina and Benjawan Kasisopa[1]

[1] MARCS Auditory Laboratories, University of Western Sydney, Sydney, Australia
{d.burnham,v.attina,b.kasisopa}@uws.edu.au

Abstract

The aim of this research is to investigate the general features of lexical tones that might contribute to their categorisation. Thai tones were presented for (a) discrimination and (b) identification by native Thai and non-native Mandarin tone language participants in auditory-only (AO), visual-only (VO) and auditory-visual (AV) conditions. Discrimination tests revealed: (i) good auditory and auditory-visual discrimination of tone pairs by Thai *and* Mandarin perceivers, (ii) significant contribution of visual information to tone discrimination in Thai *and* Mandarin perceivers; (iii) greater AV>AO augmentation at 1500 vs 500 ms interstimulus interval (ISI), showing more use of visual information for tone at phonemic (tonemic) than phonetic (tonetic) levels; and (iv) better overall discrimination – and especially large AV>AO augmentation – of contour-contour than contour-level or level-level tone pairs. Identification tests showed, as expected, that Thai participants were accurate in identifying Thai tones, using both auditory and visual information. Mandarin participants were generally able to categorize the non-native Thai tones into their native tone categories, and also used visual information, especially for contour tones. The discrimination and identification data relationship is discussed as are implications for further studies.

Index Terms: auditory-visual speech perception, cross-language studies, lexical tone.

1. Introduction

Lexical tone is the use of pitch height and contour to distinguish words in tone languages such as Mandarin (4 tones) and Thai (5 tones). Examples of these are given in Table 1, as these are the languages of interest in this paper. Tone languages comprise 70% of the world's languages [2] yet little is known about the features by which tones are discriminated and identified. Consonants are discriminated on the basis of articulatory features such as place, voicing, manner etc., and vowels on the basis of tongue height and backness, etc. For tones much less is known, and what *is* known is related to perceptual rather than articulatory features.

Table 1. *Thai & Mandarin tones showing IPA/Pinyin, tone number, verbal labels, Chao values & glosses.*

Thai	IPA	Tone No	Verbal Label	Chao Nos*	Meaning
มา	/mā:/	/ma:0/	Mid Level	33	come
หม่า	/mà:/	/ma:1/	Low Level	11	brew
ม่า	/mâ:/	/ma:2/	Falling	231	Grand-mother
ม้า	/má:/	/ma:3/	High Level	55	horse
หมา	/mǎ:/	/ma:4/	Rising	315	dog
Mandarin	**Pin-yin**	**Tone No**	**Verbal Label**	**Chao Nos***	**Meaning**
媽	mā	/ma1/	High Level	55	mother
麻	má	/ma2/	Rising	35	hemp
馬	mǎ	/ma3/	Dipping	214	horse
罵	mà	/ma4/	Falling	51	scold

*Chao numbers [1] refer to relative F0 (1=lowest, 5=highest) at critical times (onset, offset, inflection point) in production.

We know that, just as for consonants [3] and vowels [4], there is early (around 6 to 9 months) perceptual attunement for lexical tone, language specific speech perception, as a function of tone language experience [5]. Just what particular features of tone are tuned in infancy and possibly later in life are less certain. Acoustic features, phonetic and phonemic factors, and auditory-visual effects in tone perception are considered ahead of presentation of data on Thai and Mandarin participants' discrimination and identification of Thai tones.

1.1. Acoustic Features in Tone Perception

Mean fundamental frequency (F0) is a primary cue in the discrimination of Thai tones by both Thai [6] and non-native tone language and even non-tone-language [7] speakers. F0 slope is useful for Thais to discriminate level tones (those with very little pitch movement over time), and contour tones (those with tone height movement, e.g., rising of falling) in their language, and F0 direction for discriminating rising and falling Thai tones [6]. F0 onset is a good basis on which Thai tones can be discriminated [8], and this is the case for both tone and non-tone language speakers [9]. More recently, Krishnan and Gandour, electrophysiologically measuring the frequency-following response (FFR) to Thai tones, found that tone language speakers (Thai & Mandarin) have more sensitive brainstem mechanisms for representing pitch (tracking accuracy and pitch strength) than non-tone (English) language perceivers; and that tonal and non-tonal language speakers can be identified (using discriminant analysis) by their degree of response to rising (but not falling) pitches in the brainstem [10]. Krishnan and Gandour suggest that this is due to a tone-language-dependent enhancement of an existing bias towards rising (cf falling) pitch representation at the brainstem which is general across tone language speakers. To investigate general effects across tone language speakers here we investigate factors involved in Thai and Mandarin participants' perception of Thai tones.

1.2. Auditory-Visual Tone Perception

Auditory speech perception for phones is augmented [11] and modified [12] by visual speech information. Such augmentation and modification occurs in both non-tone and tone languages – Cantonese [13] and Thai (14), but it was not until 2001 that evidence for visual perception of tones was found; Cantonese participants identified Cantonese tones slightly but significantly above chance levels under certain conditions – in running speech (cf words in isolation), on monophthongs (cf diphthongs), and on contour (cf level) tones [15], and visual tone information is available both to non-native speakers of another tone language (Thai perceivers of Cantonese tones) and even non-tone language (English) speakers [16]. Two further studies have firmly established the relevance of visual cues for tone under degraded acoustic conditions: there is auditory-visual augmentation of identification of Mandarin tones by native Mandarin speakers for speech in noise but not for speech in which F0 information had been artificially removed [17], and similar results for Thai tone perception [18]. Most recently, it has been found that in Cantonese, minute rigid head movements are essential visual information for the perception of tone [19] and for Mandarin tones non-tone-language (English) perceivers use visual speech information for tone more than do native Mandarin perceivers [20]. In this study, Thai and Mandarin participants' perception of Thai tones is investigated under auditory-only, visual-only, and auditory-visual conditions.

1.3. Phonetic/Phonemic Factors in Tone Perception

For phones (consonants and vowels) there is evidence for phonemic processing in which language-specific features are perceived only by speakers of that language. This is experimentally tested by the use of different interstimulus intervals (ISIs) [21]: at 1500 ms ISI (phonemic processing) English language speakers can discriminate only native contrasts, e.g., [ba] vs [p^ha] but not the non-native Thai [ba] vs [pa]; at 500 ms ISI (phonetic processing) Thai and English speakers can discriminate the two phones, [ba] vs [pa], but not two instances of the same phone; and at 250 ms ISI (acoustic processing) two instances of the same phone ([ba_1] vs [ba_2] can be discriminated. Applied to tones, there is some evidence that native Thai speakers discriminate tones better at 1500 ms ISI (tonemic) than 500 (tonetic) 500 ms ISI, whereas non-tone (English) speakers discriminate tones better at 500 (tonetic) than 1500 (tonemic) ms ISI levels [22]. However, other studies suggest that under some conditions (i) experienced learners of Thai show better tone discrimination at 500 than 1500 ms ISI [23] and (ii) tone perception is equivalent at 500 and 1500 ms ISI [24]; and suggest that the ISI effect may be more related to short-term memory rather than phonetic/phonemic factors [23]. In this study, Thai and Mandarin participants' perception of Thai tones is investigated at each of these two ISIs.

1.4. This Study and General Hypotheses

Here Thai and Mandarin participants' discrimination (Experiment 1) and identification (Experiment 2) of Thai tones is investigated under auditory-only, visual-only, and auditory-visual conditions, and in Experiment 1 at two ISIs, 500 and 1500 ms. Generally it is expected that (i) Thais will perceive Thai tones better than will Mandarin speakers, but that Mandarin perception of Thai tones will be good; (ii) acoustic features of tones will determine both Thai and especially Mandarin perception of Thai tones; but that (iii) the addition of visual information will augment both discrimination and identification of Thai tones, especially in Mandarin perceivers. (iv) With respect to ISI (in Experiment 1), it is unclear what to expect, given previous conflicting findings [22, 23, 24], and it will be especially interesting to investigate the role of ISI in the visual and auditory-visual perception of tone.

2. General Method

Two different experiments were conducted. Experiment 1 was a paired-presentation AX discrimination task with same/different responses required and Experiment 2 was a tone identification task in which single words were required to be matched with tone-identical keywords from among tone-different distracters. The same participants were tested in each experiment. The stimuli for each experiment were recorded using the same method, as described below.

2.1. Participants

Thirty-six native Thai speakers (mean age: 28.6 years, 20 females) and 36 native Mandarin speakers (mean age: 25.2 years, 25 females) with normal hearing took part in the study. All the Mandarin speakers were naïve to the Thai language and had no knowledge of any other tone language. A questionnaire showed that none of the participants had any formal musical training longer than five years. All gave informed consent to participate in the experiment and received $30 compensation for their participation. The study was conducted under the University of Western Sydney Human Research Ethics Committee approval number H7330.

2.2. Stimuli

The stimuli were Thai syllables (/ka:/, /ki:/, /ku:/, /k^ha:/, /k^hi:/, /k^hu:/, and /fu:/) produced in citation form with the five different Thai lexical tones (3 'level' tones: low level 11, mid level 33, high level 55, and 2 'contour' tones: falling 231, and rising 315) by a native female Thai speaker (26 year-old), originally from Thailand. The speaker was required to read aloud the syllables which were displayed on a screen. The productions were audio-visually recorded in a sound-treated booth using a Lavalier AKG C417 PP microphone and a HDV Sony HVR-V1P video camera remotely controlled with Adobe Premiere software which stored the digital audiovisual recordings on a separate computer (video at 25 frames/second and 720x576 pixels; audio 48 kHz, 16 bit). Many repetitions were produced by the speaker but only three exemplars of each syllable were selected for each experiment.

The original recordings were labelled using Praat [25] and the corresponding videos were automatically cut from Praat TextGrids using a Matlab® script and Mencoder software and stored as separate video files. For each video, 200 ms were added at the boundaries to ensure that for each syllable the whole lip gesture was shown in its entirety. The sound level was normalised and all videos were compressed using the msmpeg4v2 codec. In each experiment, syllables were presented under three different modes of presentation, audio only (AO), video only (VO) and audio-visual (AV), and two audio backgrounds noise conditions: clear and noisy. For the noisy background, a multi-speaker Thai babble speech was added on top of the syllables to reach a signal-to-noise ratio of -8dB. For the AO condition, a still image of the talker was shown.

Both experiments were run using DMDX software [26] on individual Notebook Lenovo T500 computers and the sound was presented through Sennheiser HD25 II headphones at a comfortable hearing level (60 dB on average).

Prior to each experiment, the participants were given four to six trials presented in the different testing conditions to familiarise themselves with the task. The participants were given breaks in between each block and task.

3. Experiment 1: AX Discrimination

3.1. Procedure

The participants were instructed to listen to and watch a sequence of two videos of a speaker pronouncing syllables and determine whether they were same or different by pressing one of two buttons on the keyboard. For each "same" trial, different tokens of the same syllable were used, so that the task was a tone category match rather than an exact acoustic match. Participants were informed that the process was timed.

There were two between-subject factors: participants' language background (Thai vs. Mandarin Chinese) and ISI (500 ms vs. 1500 ms to test between phonetic vs. phonemic processing of tones). The within-subject factors were audio background which was also paired with consonant context (/k/ clear audio and /k^h/ noisy audio (for ½ the participants) vs. /k/ noisy audio and /k^h/ clear audio (for the other ½ of the

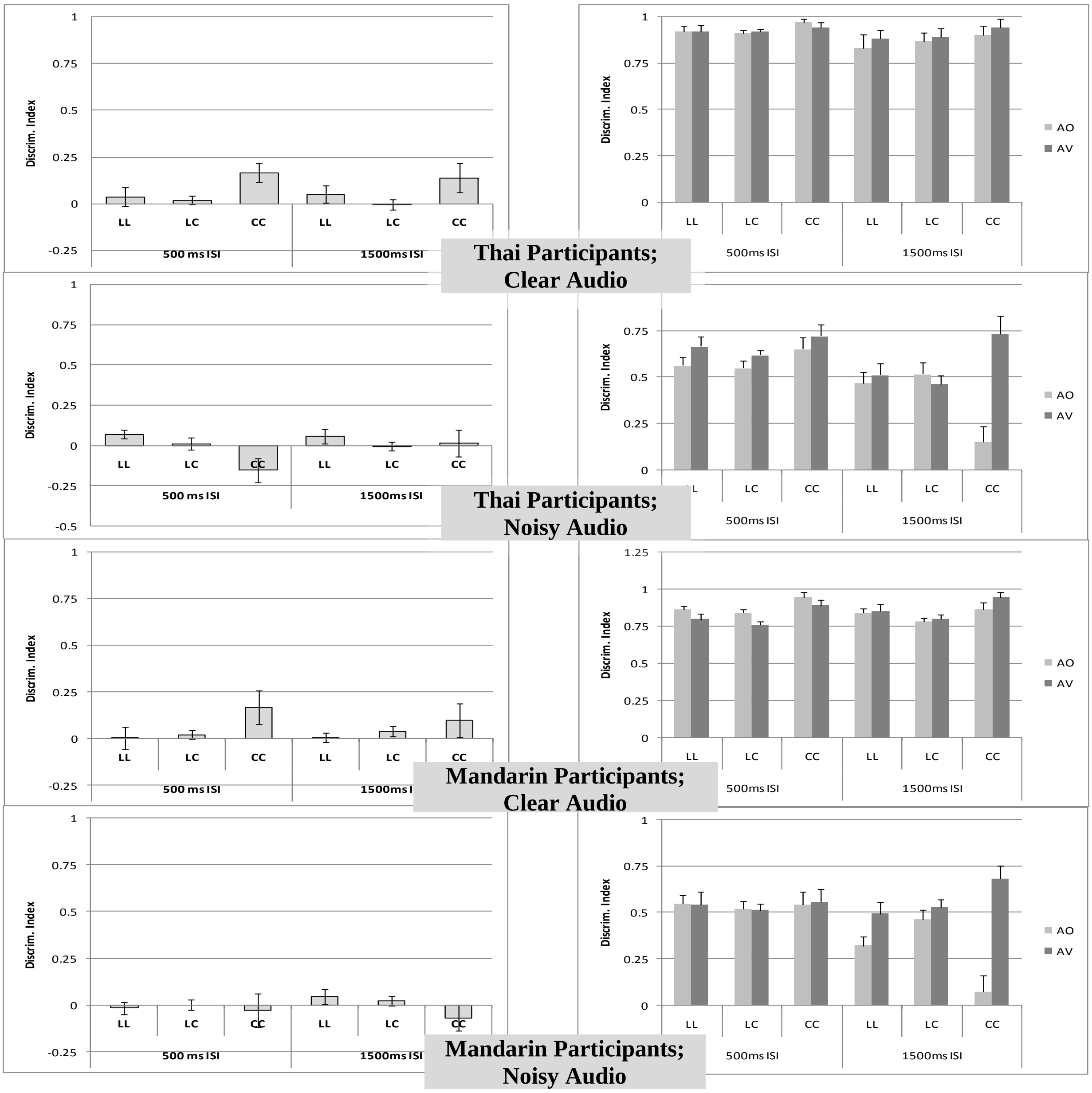

Figure 1: *Visual Only Discrimination Indices for Thai and Mandarin Participants, Clear and Noise*

Figure 2: *Auditory Only & Auditory-Visual Discrimination Indices for Thai & Mandarin Participants, Clear & Noise*

participants), mode paired with vowel context ([/a/ AV, /i/ AO, /u/ VO], [/u/ AV, /a/ AO, /i/ VO] and [/i/ AV, /u/ AO, /a/ VO]) (counterbalanced across three sub-groups of participants), and tone pair (each of the 10 paired contrasts of the 5 Thai tones). Note that Background (noise vs clear) and Mode (AO, VO, AV) were the factors of interest, with the consonant and vowel context changes simply used to introduce some degree of variability to increase generalisability and external validity of the results. So the overall design is Thai/Mandarin x 500/1500ms ISI x (noise/clear x AO, VO, AV mode x the 10 tone pairs of the 5 Thai tones) with repeated measures on the last three factors.

Clear and noisy stimuli were presented in separate blocks. The AV, AO, and VO modes were, however, mixed and randomly presented within a block to avoid any attentional bias. Each participant was tested with the 10 paired contrasts of the 5 Thai tones and the 4 AX sequences (AA, AB, BA, BB) of each tone pair and two repetitions of the entire set of trials in consecutive blocks. Therefore each participant received a total of 480 trials (=2 audio backgrounds x 3 auditory-visual modes x 10 tone pairs x 4 AX sequences x 2 repetitions).

3.2. Results

For the purposes of analyses and clearer presentation of the results, the 5 Thai tones were classified as Level (low level 11, mid level 33, high level 55), and Contour (falling 231, rising 315), so in the 10 tone pairs there were 3 level-level (LL) pairs (low-mid,11-33; low-high, 11-55; mid-high, 33-55), 6 level-contour (LC) pairs (low-falling, 11-231, low-rising, 11-315; mid-falling, 33-231, mid-rising, 33-315; high-falling, 55-231, high-rising, 55-315) and 1 contour-contour (CC) pair (falling-rising, 231-315). The accuracy results are presented as discrimination index (DI) scores focused on the “different” responses and calculated for each language background x ISI x (noise x mode), and given by [(number of Hits(number of "different" responses on different (AB or BA) trials)) MINUS (number of False Positives (number of "different" responses on same (AA or BB trials)] / number of trials for that pair (=4). A score of 1 shows perfect discrimination, zero chance responding, and -1 or scores significantly below zero would indicate perverse responding. DI scores for VO trials and for AV and AO trials are shown in Figures 1 and 2 respectively.

DI scores were analysed separately for (i) VO scores and for (ii) AO and AV scores together in similar analyses of variance (ANOVA) with the design Language Background (Thai, Mandarin) x ISI (500, 1500) x (Noise (clear, noise) x Tone Contrast Type (LL, LC, CC)) with repeated measures on the last 2 factors. For the AO/AV analysis AO vs. AV was added as an additional within-subjects factor. For all analyses α was set at 0.05.

VO scores are shown in Figure 1. The ANOVA revealed that there were no significant between-subjects effects (Language Background, ISI, and their interaction). Within-subjects there was understandably no main effect of clear vs. noise. However, there was one significant of Noise/Clear x Tone Contrast Type – CC vs. (LC+LL), $F(1,68) = 9.58$. This shows that all participants – across language groups and ISIs – Contour-Contour tone contrasts were more easily discriminated than Level-Contour of Level-Level contrasts but only in the clear audio condition. It appears that, as has been found in previous studies, although only for non-tone language participants [19], the presence of *audio* noise affects *visual* perception of lexical tone. This result is borne out by t-tests against chance that were conducted in order to ascertain if any of the VO responses were significantly above chance. These revealed that Thai participants discriminated Contour-Contour contrasts in Clear audio significantly above chance at 500 ms ISI, $t(17) = 3.37$, and at 1500 ms ISI, $t(17) = 1.76$; and that Mandarin participants discriminated Contour-Contour contrasts significantly above chance at 500 ms ISI, $t(17) = 1.84$, but not at 1500 ms ISI, $t(17) = 1.07$. No other tests against chance were significant either in clear or noisy audio.

AO and AV scores are shown in Figure 2. The ANOVA revealed that Thai participants (M=0.73) were understandably better than Mandarin participants, $F(1,68) = 8.55$, although Mandarin participants were still quite proficient (M=0.66) at this non-native tone task. The main purpose of this study was to investigate visual perception of tone – whether there is any augmentation of tone discrimination when visual information is added, i.e., in AV vs. AO. Discrimination indices were indeed higher for AV (M=0.73) than AO (M=0.66), $F(1,68) = 4.86$. The locus of the greatest AV>AO effects are of particular interest. AV>AO augmentation was greater at 1500 ($M_{1500} = 0.73$; $M_{500} = 0.59$) than 500 ($M_{1500} = 0.74$; $M_{500} = 0.74$) ms ISI, $F(1,68) = 4.51$, and this effect was ever so slightly greater for Mandarin participants, $F(1,68) = 5.87$. Thus, it appears that phonemic (tonemic) information is more relevant to visual tone discrimination than tonetic information, and non-native Mandarin participants are especially good at picking up this information to augment their AO tone discrimination. While these results are true for all conditions, noise understandably enhances the AV>AO effect, $F(1,68) = 7.42$. Turning to tone types, discrimination is generally better across all presentation modes (AO, VO, AV) for contour-contour tone contrasts than for level-level tone contrasts or level-contour tone contrasts (CC>(LL+LC)), $F(1,68) = 5.28$, especially at 1500 vs 500 ms ISI, $F(1,68) = 7.86$; and there is especially large AV>AO augmentation for CC contrasts for Mandarin as opposed to Thai perceivers, $F(1,68) = 5.50$; and especially large AV>AO augmentation for CC contrasts at 1500ms ISI, $F(1,68) = 32.99$, particularly in noise, $F(1,68) = 17.99$ (see Figure 2).

3.3. Discussion

Thai perceivers perceive Thai tone better than do non-native Mandarin perceivers, but their superiority is not great; there is quite some cross-tone-language perception of tone. Over and above this slight superiority, there is across the board evidence for the visual discrimination of visual tone. This is especially the case for contour-contour tone contrasts which provide more visual information than do level-level tone contrasts or level-contour tone contrasts, a fact evident in both the VO conditions and in visual augmentation of AO by V in the AV condition. AV>AO augmentation is particularly strong for both Thai and Mandarin perceivers, but significantly more so for the non-native Mandarin speakers; and specifically present for both Thai and Mandarin perceivers at 1500 rather than 500 ms ISI, but again, significantly more so for the non-native Mandarin speakers. It appears that there is some tonemic tone- language-general information (well, general across Thai and Mandarin so far) that is particularly useful for the visual perception of tone, and while this is used by native and non-native tone language speakers alike, non-native speakers appear to seek this out to a greater extent.

4. Experiment 2: Tone Identification

4.1. Tone identification predictions

Experiment 2 was designed to assess how native and non-native tone speakers identify tones. Native speakers should correctly identify each tone category and their performance should be better in AV than in AO. For non-native Mandarin speakers, it is interesting to know how they will categorise the five Thai tones, and the influence of their native Mandarin tone categories.

Table 2. *Correlations between Thai F0 tone & Mandarin F0 tone normalised values.*

	Mand55	Mand35	Mand214	Mand51
Thai11	0.03	-0.764*	-0.569*	**0.91***
Thai33	-0.198	-0.83*	-0.813*	**0.916***
Thai55	-0.357	**0.26**	0.067	-0.127
Thai315	0.215	**0.947***	**0.952***	-0.901*
Thai231	-0.23	-0.94*	-0.917*	**0.955***

(* p<.01; **bold**: = highest positive correlation)

As F0 variations are the main acoustic features for tones, we can reasonably base predictions on F0 shape similarities between the Thai and the Mandarin tones. We calculated correlation coefficients between time-normalized values of Thai F0 tone and Mandarin F0 tone over time for each tone pair (see Table 2). The original F0 values were extracted from a corpus of /fu:/ syllables (part of another project) pronounced with the 5 Thai tones and repeated at least 5 times each by 4 native Thai female speakers, and Mandarin /fu/ syllable productions with the 4 tones of Mandarin of 3 native Mandarin female speakers. From each Thai-Mandarin tone pair, we consider only the highest positive correlation. Thus predictions for Thai tone categorisation by Mandarin speakers are:

- Thai tone 11 will be perceived as Mandarin tone 51;
- Thai tone 33 will be perceived as Mandarin tone 51;
- Thai tone 55 will be perceived as Mandarin tone 35. Note, none significant, but highest +ve value is for 35.
- Thai tone 315 will be perceived as either Mandarin tone 214 or 35 (as there are two similar high correlations);
- Thai tone 231 will be perceived as Mandarin tone 51.

4.2. Procedure

Participants were instructed to listen to and watch a video of a syllable pronounced by the speaker and match it by mouse click with one of the given keywords in their own language displayed on the screen. The Thai syllable /fu:/ with the five Thai tones was used for the identification task. The 5 keywords for Thai were: ฟู (fu33), ฝู่ (fu11), ฟู่ (fu231), ฟู้ (fu55) and ฝู (fu315), and the 4 keywords for Mandarin were: 夫 (fu55), 扶 (fu35), 抚 (fu214) and 富 (fu51). In addition, the category “unknown” was given as a possible answer when participants were unable to select one of the provided keywords. The identification task consisted in 180 randomised trials (5 tones x 3 modes [AV, AO, VO] x 2 backgrounds [clear vs. noise] x 3 exemplars x 2 repetitions) presented in two separate blocks, clear speech and speech in noise.

4.3. Results

Audio (AO & AV scores) and visual-only (VO) data were analysed separately. Only the AO and AV results are reported here. To examine Thai participants’ tone identification and the effect of visual augmentation and noise type, five 2x2x5, Background (clear/noisy) x Mode (AO/AV) x Tone (Thai tones 33, 11, 231, 55, 315), ANOVAs were conducted, with planned contrasts on the tone factor, predicting that identifications of the presented tone being significantly more frequent than the other 4 tones combined. For the Mandarin participants five 2x2x4, Background (clear/noisy) x Mode (AO/AV) x Tone (Mandarin tones 55, 35, 214, 51), ANOVAs were conducted, with planned contrasts on the tone factor based on the F0 predictions (see Table 2) such that identifications for the predicted tone would be significantly greater than for the other 3 tones combined (or in the case of Thai tone 315, that identifications for Mandarin tones 35 and 214 would be significantly greater than for the other two, 55 and 51). α was set at 0.05 throughout. Thai and Mandarin ANOVA results are shown in Tables 2 and 3 respectively.

Table 3 shows that the five Thai tones were in general correctly identified by Thai participants with overall percent correct ranging from 68.8% (Tone11) to 94.7% (Tone33). There was generally better performance in clear than in noise, in AV than in AO, and generally greater AV>AO augmentation in noise than in clear. The exception was tone 55 for which there was good identification across all conditions (91.6%). The two other level tones (11 & 33) were better identified in clear than in noise; the contour tone 315 was better identified in AV; and the other contour tone, 231,was better identified in AV, but only in noise.

Mandarin participants (see Table 4), were able to categorise the Thai tones as a function of their Mandarin tone system. For 3 of the 5 Thai tones (11, 315 & 231) their overall ability to use one of the 4 Mandarin tone categories (and not fail to respond, or respond ‘unknown’) was facilitated by the use of visual information, especially in noise; and it is notable that 2 of these 3 are the contour tones; contour tones appear to convey some wealth of visual speech information [15]. The main exception to the AV>AO augmentation was Thai tone 55, for which performance was quite singular (a mean of 90.9% responses as Mandarin tone 35), and this tone was the one for which Thais also performed uniformly well (91.6%).

4.4. Discussion

Predictions based on F0 variations were upheld for only two Thai tones, tones 55 and 315. For the other three Thai tones, it was predicted that they would be perceived as Mandarin tone 51, but Thai tones 33 and 231 were instead perceived as Mandarin tone 55; and for Thai tone 11, the pattern of identification was complex, with Mandarin tone 55 responses in clear and both Mandarin tones 214 and 51 responses in noise. One explanation for this poor set of predictions is that they were based on shape similarity across languages using

Table 3: ***Thai Participants’*** *Identification of Thai tones: Data & ANOVAs. ANOVA cell values are Fs for df = 1,35;* F_{crit} *=4.121*

Thai Tone	Data %Corr	ANOVA Results						Comments
		Augmentation in:			Prediction			
		Clear	*AV*	*Clear x AV Worst = AO/Noise?*	*Upheld?*	*> in Clear?*	*> in AV?*	
11	68.8	24.2	22.8	21.4	409.9	48.5	NS	Better prediction in Clear than Noise
33	94.7	4.9	4.9	4.9	3421.9	14.9	NS	Better prediction in Clear than Noise
55	91.6	NS	NS	NS	537.9	NS	NS	Good prediction in all conditions
315	82.4	7.3	6.0	5.4	257.1	NS	8.80	Better prediction in AV
231	76.5	7.8	6.4	6.4	197.5	14.0	14.0 Only in Noise	Better prediction in Noise AV

Table 4:***Mandarin Participants’*** *Identification of Thai tones: Data & ANOVAs. ANOVA cell values= Fs for df = 1,35;*F_{crit} *=4.121*

Thai Tone	Data		ANOVA Results						Comments
			Augmentation in:			Prediction			
	Pred **%**	*Actual* **%**	*Clear*	*AV*	*Clear x AV Worst = AO/Noise?*	*Upheld?*	*> in Clear?*	*> in AV?*	
11	25.8	33.0	26.7	18.2	17.6	NS	Clear x AV 6.03; Best prediction in AV + Noise		**T11 → M51 prediction** not upheld; Clear: T11→M55, F=31.6; Noise: T11→M214 & M51, F=40.2
33	2.6	86.8			10.0	265.4			**T33 → M51 prediction** not upheld Thai33→Mand55 F=791.9
55	90.9	90.9	**N>C** 4.13			753.7	**N>C** 7.3		**T55 → M35 prediction** upheld & more so in noise.
315	59.6 34.5	59.6 34.5	6.1	8.8	6.7	3273.2 10.1			**T315 → M214 prediction** upheld **T315 → M35 prediction** upheld
231	10.3	73.7		11.7	8.2	28.7			**T231 → M51 prediction** not upheld T231 → M55, F=296.0

Grey shading shows effects that are significant at p<.05 or more, with the significant F-values shown. T = Thai, M = Mandarin time-normalised F0 values. Examination of the original F0 values shows that Mandarin tone 51 is clearly shorter in duration than the other tones. So one reason why our predictions for Thai tones 33, 11 and 231 were not upheld could be that even though the F0 shape were similar, tone duration was a more important basis for Mandarin participants' decision-making. Another reason is that visual information was not taken into account in the predictions and for 2 (11, 231) of these 3 Thai tones there was indeed augmentation in decision-making in AV compared with AO.

More generally it can be concluded that Mandarin perceivers are able to use their own tone system to categorise tones; that both Thai and Mandarin use visual information in categorising tones; that contour tones uniformly appear to contain rich visual information; and that future predictions of tone category membership must take into account more acoustic variable than just F0, and visual information as well.

5. General Discussion and Conclusions

Most generally, the results show that tone is perceived well across tone languages; Mandarin perceivers are adept at discriminating Thai tones (Experiment 1), and can assign Thai tones to Mandarin tone categories quite well (Experiment 2). At a general level this shows that tone features can be transferred across tone languages. In addition, it is notable that it is not just acoustic, but also visual speech information is used both to discriminate and identify tones by both native and non-native tone language speakers. Moreover, as found by [20] it is the non-native speakers who more adeptly seek out visual information for tone. For both language groups, the richest source of AV information is for contour-contour contrasts in Experiment 1, and contour tones in Experiment 2, showing that there is something pervasive about visual information in contour tones [see also15]. In addition, in Experiment 1 it was found that both Thai and Mandarin participants show greater AV>AO augmentation at 1500 ms ISI. On the assumption that this reveals something about tonemic processing [but see 23], it appears that it there may be tonemic information that is general across a number of tone languages, and that, just as Krishnan and Gandour [10], suggest, there may be enhancement of an existing bias towards the representation of pitch contours (rising more than falling in their formulation) at the brainstem, which is general across tone language speakers. Just how general this is, e.g., is there some tone features that are general across some languages moiré than others; whether this generality is the case also for visual speech; and whether such auditory and visual tone information is of a tonemic rather than a tonetic nature, are questions that await resolution in future research.

6. Acknowledgements

We appreciate Mr. Leo Chong's data collection assistance; Drs Karen Mattock & Connie So's helpful input; & Australian Research Council funding (DP0988201) to the 1st author.

7. References

[1] Chao, Y-R., A system of tone-letters, Le Maître Phonétique, 45: 24–27, 1930.

[2] Yip, M. Tone. Cambridge, UK, Cambridge Univ. Press, 2002.

[3] Werker, J. F. & Tees, R. C. The organisation & reorganisation of human speech perception. AnnRev Neurosci., 15, 377-402, 1992.

[4] Polka, L. & Werker, J. F. Developmental changes in perception of nonnative vowel contrasts. J. Exp. Psych: HPP, 20, 421-35, 1994.

[5] Mattock, K., & Burnham, D. Chinese and English infants' tone perception: Evidence for perceptual reorganization. Infancy, 10, 241-265, 2006.

[6] Gandour, J. (1979). Perceptual dimensions of tone: Thai. In G.I. Liem (Ed) South-East Asian Linguistic Studies Vol 3). ANU, Canberra: Dept. of Ling. (Pacific Linguistics, Series C, No.45).

[7] Gandour, J. (1983). Tone perception in Far Eastern languages. J. Phonetics, 11, 149-175.

[8] Saravari, C. & Imai, S. Perception of tone and short-long judgements of vowel of variants of a Thai monosyllabic sound. J. Phonetics, 11, 231-242, 1983.

[9] Burnham, D. K., Kirkwood, K., Luksaneeyanawin, S. & Pansottee, S. Perception of Central Thai tones & segments by Thai & Australian adults. Pan-Asiatic Linguistics: Proc.3rd Internat. Symp. Language & Linguistics. Bangkok: Chulalongkorn Univ. Press, 546-560, 1992.

[10] Krishnan, A., Gandour,J.T., Bidelman, G.M. The effects of tone language experience on pitch processing in the brainstem. J Neurolinguistics, 23, 81–95, 2010.

[11] Sumby, W. H. & Pollack, I. Visual contribution to speech intelligibility in noise. JASA, 26, 212-215, 1954.

[12] McGurk, H., and McDonald, J. Hearing lips and seeing voices. Nature 264, 746-748, 1976.

[13] de Gelder, B., Bertelson, P., Vroomen, J., and Chen, H. C. Interlanguage differences in the McGurk effect for Dutch and Cantonese listeners," in *4th European Conference on Speech Comm. & Technology*, (Madrid, ISCA), 1699-1702, 1995.

[14] Burnham, D. K. Auditory-visual perception of Thai consonants by Thai and Australian listeners. Pan-Asiatic Linguistics: Proc.3rd Internat. Symp. Language & Linguistics. Bangkok: Chulalongkorn Univ. Press, 531-545, 1992.

[15] Burnham, D., Ciocca, V. & Stokes, S. Auditory-visual perception of lexical tone. Eurospeech, Eds. B.L.P. Dalsgaard, & H.Benner (Aalsborg, , ISCA, Bonn, Germany), 395-8, 2001.

[16] Burnham, D., Lau, S., Tam, H., and Schoknecht, C. Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers. AVSP, Aalborg, Denmark, D. Massaro, Light, J., & Geraci, K. (Eds). Causal Productions, Adelaide, South Australia, 155-160, 2001.

[17] Mixdorff, H., Hu, Y. & Burnham, D. Visual cues in Mandarin tone perception. 9th Eurospeech, ISCA, Bonn, 405-8, 2005.

[18] Mixdorff, H., Charnvivit, P. & Burnham, D. Auditory-visual perception of syllabic tones in Thai," AVSP, E. Vatikiotis-Bateson, D. Burnham, D., S. Fels (Eds) Causal, Adelaide, British Columbia, Canada), 3-8, 2005.

[19] Burnham, D., Vatikiotis-Bateson, E., Yehia, H., Ciocca, V., Haszard Morris, R., Hill, H., Vignali, G., Reynolds, J., Jones, C., Tam, H., Bollwerk, S. & Reid, A. Visual perception and production of lexical tone: the role of head and face motion.

[20] Smith, D., & Burnham, D. (submitted) Facilitation of Mandarin tone perception by visual speech in clear and degraded audio: Implications for cochlear implants. Accepted with revisions.

[21] Werker, J. F. & Logan, J. Cross-language evidence for three factors in speech perception. Perc. & Psych, 37, 35-44, 1985.

[22] Burnham, D. & Francis, E. The role of linguistic experience in the perception of Thai tones. In A. S. Abramson (Ed) SouthEast Asian Linguistic Studies in Honour of Vichin Panupong (Sci. of Lang, v8. Bangkok: Chulalongkorn Univ. Press, 29-47, 1997.

[23] Wayland, R. & Guion, S. Training English and Chinese listeners to perceive Thai tones: A preliminary report. Lang. Learning, 24: 113-129, 2004.

[24] Wayland, R. & Guion, S. Perceptual discrimination of Thai tones by naive and experienced learners of Thai. Applied Psycholinguistics, 24, 113-129, 2003.

[25] Boersma, P., Praat, a system for doing phonetics by computer.Glot International, 5(9/10):341-345, 2001.

[26] Forster, K.I., & Forster, J.C., DMDX: a windows display program with millisecond accuracy, Behavior Research Methods, Instruments, & Computers, 35(1):116-124, 2003.