Perception of Tone by Native Mandarin Chinese Listeners: Optimal Auditory Perception*

Andrew C.-J. Hung[a] and Rong-fu Chung[b]
[a]*National Cheng-Kung University, Taiwan*
[b] *Southern Taiwan University of Science & Technology, Taiwan*

1 Introduction

In Mandarin Chinese, a syllable template is CGVE, where C is onset, G medial glide, V nucleus vowel, and E a nasal consonant or a glide. However, V is the only indispensable segment. In addition, every syllable gets a tone, which differentiates meanings. For instance, the Mandarin Chinese syllable [bi] has the meaning "pen" when produced with a dipping pitch contour, but the meaning of "nose" when produced with a high rising contour. Therefore, perception of a lexical tone is involved with processing of acoustic signals in phonetics and decoding process in lexical or semantic functions. Accordingly, it is difficult to single out which effect matters in the studies of tone perception. In other words, neither purely acoustic nor purely phonemic-based model can provide crucial evidence. In this article, we try to sort out effects ascribed to the acoustics and the phonetics of the speech signals in the absence of lexical-semantic processing, or other intervening processes.

One of our methods is to use a sine-wave speech (for details see *Stimuli* section below), a technique for synthesizing speech by replacing the formants (main bands of energy) with pure tone whistles. A sine-wave speech is perceived either as speech or non-speech, depending on listeners' previous language experience (Möttönen et al., 2006; Remez et al., 1981). Behavioral studies (e.g., Remez et al., 1981; Whalen & Liberman, 1987) reveal that acoustic signals could be perceived categorically or not depending on whether it forms part of a sound complex perceived as a syllable or not . For example, in Remez et al.'s study (1981), naïve subjects failed to perceive sin-wave stimuli as speech. However, when the subjects were instructed about the speech-like nature of sine-wave stimuli, they could easily understand it. Neural studies also found more marked activation in the cortical regions of the left hemisphere was seen only in those subjects who were able to identify the sine-wave stimuli as speech (Möttönen et al., 2006). These findings indicate that identical physical sounds are processed categorically are associated with whether they are perceived as speech or not. This characteristic makes sine-wave speech a feasible medium to investigating tone perception.

Another method we use is lexical gap. Out of the vowel repertoire in Mandarin Chinese, we chose vowel /a/ as the tone carrier. Vowel /a/ in Mandarin Chinese carries few lexical meanings, except for some emotional and affective connotations, such as questioning, interjection, exclamation, and annoyance. On occasions, it serves as a prefix to denote familial relationship in the Southern Min language. Furthermore, /a/ with a dipping tonal contour maps to no lexeme in Mandarin Chinese speech, and thus results in an accidental gap. Because of the accidental gap, no lexical meaning is anticipated when the signal is heard. With lexical gap, acoustic and phonetic processing can be better observed without the interference caused by lexical processing.

We integrated these two methods to address the afore-mentioned issues. We examined participants' behavioral responses to stimuli of sine-wave tone (SW), as compared to those of lexical tone (LX). The stimuli were presented in a delayed-match-to-sample paradigm (see Figure 1), in which pairs of tone stimuli were played interspersed with a delay and the participants responded after hearing the second stimulus. The pair-wise tone stimuli either comprised two familiar tone contours (FF), or one invented novel contour and one familiar tone contour (NF). In the experiment, the participants performed two tasks. In one, they were required to judge whether the pair-wise tone stimuli were identical or not, and in the other, they were required to judge whether the contour of the pair-wise tone stimuli fitted into the same type of contour. In the present article, the first task is referred to as the auditory discrimination (AUD), and the second task is referred to as the category discrimination (CAT).The participants' performance, in terms of reaction latency, accuracy, and

* We thank the audience who came to the poster presentation of the current study at Phonology 2014 held at Massachusetts Institute of Technology (MIT) on Sept. 19-21, 2014.

Proceedings of AMP 2014
Completed March 1, 2015

efficiency were analyzed to determine in which type of presentation mode—SW or LX, in which type of context—FF or NF, they were able to discriminate the pair-wise tone stimuli with greater ease.

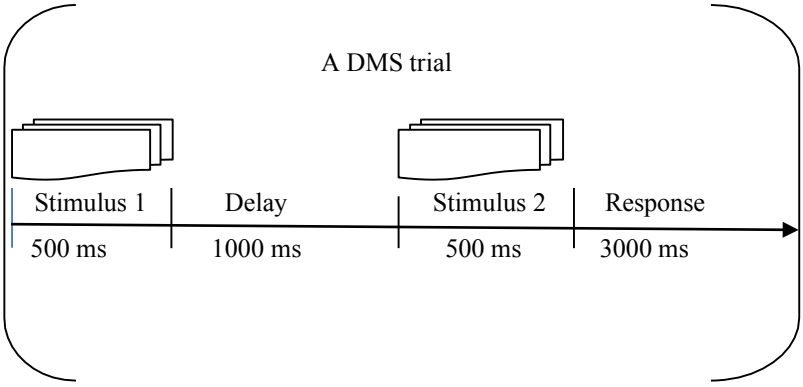

Figure 1. Delayed-match-to-sample task (DMS).

2 Materials and Methods

2.1 *Participants*

Twenty-two native listeners of Mandarin Chinese male teenagers, aged around 17, right-handed, with normal hearing ability, participated in the study. The participants gave their informed consent to the study. They were paid for the participation. The participants gave their informed consent to the study. Each participant filled out a language-background questionnaire before beginning the experiment.

2.2 *Stimuli*

In the present experiment, we employed two types of *Sound*: sine-wave tone and lexical tone. Lexical tones refer to tones that accompanied vowels. Sine-wave tones refer to tones that were devoid of vocalic information (i.e., formats) but still kept the same pitch contour as their counterparts of lexical tones.

Sine-wave tone stimuli were generated with the help of *Praat*[1] (Boersma & Weenink, 2013). An example of /a/ was downloaded from a Mandarin Chinese Audio Pronunciation Guide website, (Qiu, 2012). The vowel /a/ was used as a prototype, from which all the auditory stimuli derived. The prototypical /a/ with distinctive lexical tone was manipulated by extracting its pitch tier and then resynthesizing it into an auditory sine-wave signal. The resynthesized sine-wave tone retained the same pitch contour as the prototype, but the formats were replaced with sine-wave whistles. The spectrograms of the tone stimuli of dipping contours and up-down contours are illustrated in Figure 3 and Figure 4 for examples: sine-wave tone (left) vs. lexical tone (right). Creation of the tone stimuli is given in detail below.

(a) Dipping contours (Figure 2): The dipping contours (down-up sweeps) were similar to those of the third tone of Chinese Mandarin. They were referred to as Tone 3 in the current study. The first exemplar of the dipping tone contour had a turning point at the one third of the duration, and at the frequency of 50 Hz. The dipping contours started at 150 Hz, and ended at 300 Hz. And the other four exemplars were generated by increasing the starting and the ending frequencies, and the frequency of the turning point at a step of 20 Hz respectively. The dipping tone contours were at frequencies of 150-50-300,170-70-320,190-90-340, 210-110-360, and 230-130-380.

(b) Up-down sweeps (Figure 3): Up-down sweeps were not existent in the tone repertoire of Mandarin Chinese. They were referred to as Tone 5 in the present study. The five up-down tonal contours had onset frequencies at 300, 320, 340, 360, 380 Hz, and ending frequencies at 150, 170, 190, 210, and 230 Hz, and all sweeps had a turning point, which fell at one third of the duration, and at the frequency of 50, 70, 90, 110, and 130 Hz. The up-down pitch contours were at frequencies of 300-50-150, 320-70-170, 340-90-190, 360-110-210, and 380-130-230 Hz respectively.

[1] *Praat* is a free computer software for speech signal editing, labeling, and synthesizing.

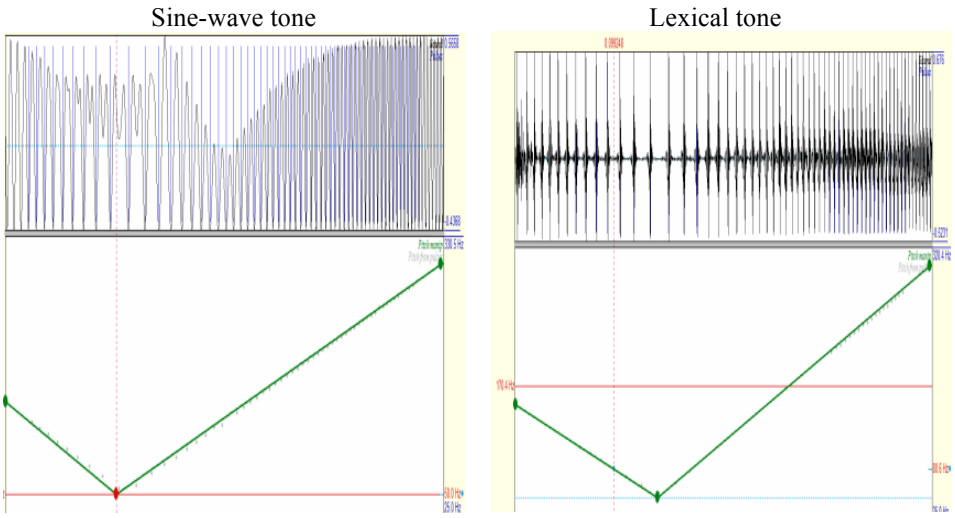

Figure 2. Tone 2: Dipping contours.

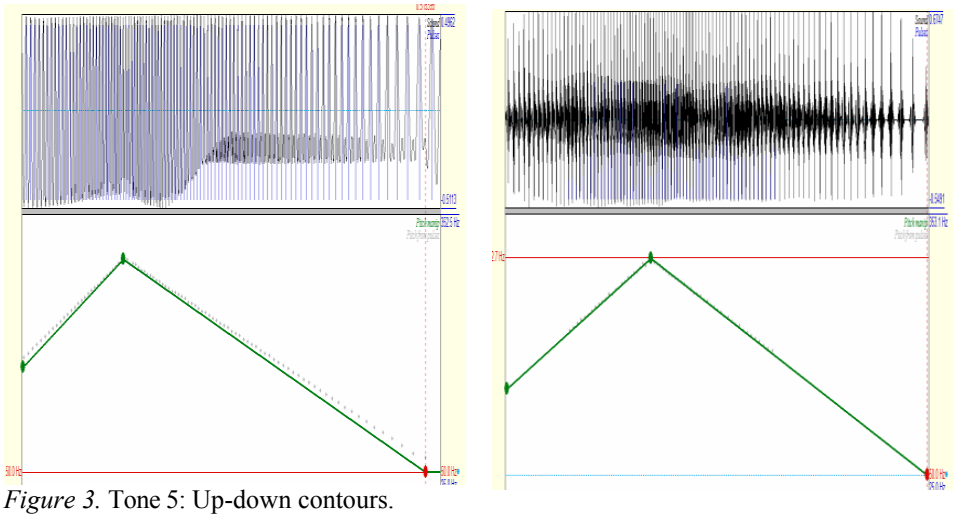

Figure 3. Tone 5: Up-down contours.

In the present experiment, the stimuli were presented in a delayed-match-to-sample (DMS) paradigm (see Figure 1). The pair-wise tone stimuli comprised either two familiar contours (FF) or the invented novel contour and one familiar contour (NF). The constructs of the pair-wise stimuli are illustrated in Figure 4. FF represents a context where pair-wise tone stimuli consisted of exemplar of Tone 3 and another exemplar of tone contours in Mandarin Chinese, and NF refers to a context where pair-wise tone stimuli were made up of one exemplar of an invented tone contour, which we named Tone 5 and another exemplar of tone contours in the tone repertoire of Mandarin Chinese, excluding Tone 3.

a. Familiar-Familiar Context

Tone 3	Tone 1
	Tone 2
	Tone 3
	Tone 4

b. Novel-Familiar Context

Tone 5	Tone 1
	Tone 2
	Tone 4
	Tone 5

Figure 4. The constructs of the pair-wise tone stimuli

2.3 *Procedure*

The participants were tested individually wearing KHM-7602 headphones in a language lab. Before the test, the participants were instructed how to respond, and had a three minutes' practice to familiarize themselves with the procedure. For the discrimination tests, the native Mandarin Chinese listeners heard pairwise tone tokens, and then they needed to respond as quickly and accurately as possible. The decision was made according to either of the two criteria: (a) whether the pairwise tone stimuli were identical, and (b) whether the pairwise tone stimuli fitted into the same category of tone contour. The participants denoted their judgment by making one of two right finger key presses. The left bottom (the index finger) indicated "NO", while the right button (the middle finger) indicated "Yes". The response time was fixed, regardless of the time taken by the participants to indicate the judgment. The response latency and accuracy of discriminating each pairwise token was recorded by DMDX (Forster & Forster, 2003)[2].

Each trial lasted for 5000 ms, with two stimuli of 500 ms each, an inter-stimulus interval of 1000 ms, and 3000 ms of response time (see Figure 1). Each test block consisted of ten trials and thus lasted for 50 seconds. A run of the tone discrimination test consisted of four test blocks and four rest blocks. A run of test altogether lasted approximately 330 seconds. The experiment had two runs of discrimination tests, one in the mode of sine-wave tone, and the other in the mode of lexical tone. Each run consisted of two auditory discrimination tasks, and two category discrimination task. The instruction for the auditory discrimination task was "Are they identical?" and the instruction for the category discrimination task was "Do they fit into to the same category of tone contour?" The visual instruction was presented in Chinese. Visual instruction was given for 30 seconds in each rest block, and disappeared as soon as the testing started. The order of the tasks was randomized. Two runs of discrimination tests took about eleven minutes to complete. The timeline of one run of the discrimination test is illustrated in Figure 5. The presentation of stimuli was automatically mediated by DMDX, which ensures millisecond timing accuracy. The program was installed on an HP notebook, with an i7-3610QM CPU at 2.30GHz.

Rest	Test 1	Rest	Test 2	Rest	Test 3	Rest	Test 4
	AUD*x*FF		AUD*x*NF		CAT*x*FF		CAT*x*NF
+ VI	10 trials	VI	10 trials	VI	10 trials	VI	10 trials
10 + 30s	50s	30s	50s	30s	50s	30s	50 s

Figure 5. Timeline of one run of the discrimination test. VI = Visual Instruction. The order of the test blocks was randomized.

The distribution of the trials in each task block of the pair-wise trials went as follows. For the discrimination task, four out of the ten pairwise trials were identical, and six pairs were not. For the category discrimination task, seven of the ten pairwise trials belonged to the same category of contour, and the remaining three pairwise trials did not. To guarantee a random distribution throughout the experiment, the scrambling function of DMAX was employed. The function first randomly ordered the trials in each block, and then randomly ordered the blocks. Furthermore, scrambling ensured that no two participants received the same sequence of items. Besides, the pairwise tone tokens were counterbalanced.

Because sound input from one ear projects to the contralateral hemisphere, and because processing of lexical tone has been observed lateralized in the left hemisphere of the brain (e.g., Gandour et al., 2003; Klein et al., 2001; Zatorre, 2003), when tone tokens are presented through the right ear only, they are projected to the left hemisphere (Frost el al., 1999; Tervaniemi & Hugdahl, 2003). Accordingly, in the current experiment, tone stimuli were presented through the right ear only with a view to observing the participants' behavioral responses in discriminating the pair-wise of tone stimuli.

[2] DMDX is a script interpreting system for screen control, stimulus presentation, and timing for cognitive experiments. Retrieved from http://www.indiana.edu/~clcl/Q550_WWW/DMDX.htm.

2.4 *Data analysis*

Three-way factorial analysis of variance (ANOVA) was conducted by following the GLM univariate procedure (SP*SS 18*), which provides analysis of variance for one dependent variable by one or more factors or variables. We examined the effects of the independent variables on the dependent variable of discrimination latencies, and the interaction between the independent variables. Post-comparisons within and between levels of each independent variable were further conducted. In addition, we proposed *Optimal Auditory Perception* to account for the superiority of one condition over the other.

3 Results and Discussion

In total, we collected 1760 responses. Response latencies below 200 ms or at the cut-off values 3000 ms were removed from the analysis. There were 46 responses whose latencies were less than 200 ms, and 11 responses whose latencies exceeded 3000 ms. The percentage of invalid response was 3.3%. Among the 1702 eligible responses, there were 384 incorrect responses (21.8%), and 1318 correct responses (about 74.9%). Only the latencies of the correct responses were further analyzed. Response latency, accuracy[3], and efficiency[4] of discriminating these pairwise tone stimuli were compared within and between levels of each independent variable: *Task (AUD vs. CAT)*, *Sound (SW vs. LX)*, and *Context (FF vs. NF)*.

In the following analysis, the main effects of each independent variable are presented first, and then post-comparisons within and between levels of each independent variable are given. Implications of the results come at the end of the analysis of each independent variable.

3.1 *Complexity of Task*

3.1.1 *Performance between AUD and CAT*[5] The results of ANOVA reveal that the mean latency for AUD and that for CAT was significantly different, $F(1, 1310) = 5.61$, $p < .05$, with the mean latency for AUD (M = 724 ms, SD = 345 ms) shorter than that for CAT (M = 770 ms, SD =416 ms). In addition, the accuracy rate for AUD (84%) was also better than that for CAT (71%). When accuracy is viewed in relation to response latency, that is, in terms of efficiency, the participants performed more efficiently for AUD (*eff.* = 0.98) than they did for CAT (*eff.* = 0.79).

AUD required the participants to judge whether the pairwise tone tokens were identical or not. The task could be achieved by referring to acoustic information of the stimuli, without involving categorical knowledge. In contrast, CAT, where the participants were instructed to judge whether the pairwise tone stimuli fitted into the same category of contour or not, involved not only acoustic information of the pairwise tone tokens but also categorical knowledge of these tone stimuli. Probably due to the extra cognitive demand, CAT tended to have longer response latencies than AUD.

3.1.2 *Efficiency in Context x Sound within and between AUD and CAT* Of the four conditions for AUD, the discrimination efficiency in NF *x* SW *x* AUD (*eff.* = 0.27) was the best. We infer that the novelty of the invented tone contour and the simplicity of the sine-wave tone working together leading to the best discrimination efficiency. In contrast, of the four conditions for CAT, the discrimination efficiency in FF *x* LX *x* CAT was the worst (*eff.* = 0.15). It is probable that the complexity of the lexical tone stimuli contributed to the result in spite of the familiarity of the tone contours to the participants. Among all the conditions of AUD and CAT, the discrimination efficiency in NF *x* SW *x* AUD is the optimum.

---

[3] Accuracy rate was calculated by the number of correct responses in relation to all responses.

[4] Efficiency was calculated by the number of correct responses in relation to time expenditure. The abbreviation e*ff.* in the parenthesis *s*tands for efficiency in the paper.

[5] For brevity, abbreviations are used in the following sections of the present article. The abbreviations go as follows: AUD = the auditory discrimination task, CAT = the category discrimination task; SW = sine-wave tone stimuli, LX = lexical tone stimuli; FF = pair-wise stimuli comprising two familiar contours, NF = pair-wise stimuli comprising the invented contour and one familiar contour. For example, NF x SW x AUD means the auditory discrimination (AUD) of the novel-familiar pair-wise tone stimuli (NF) in the mode of sine wave tone (SW).

3.1.3 *Implications--Sound x Context: AUD vs. CAT* The difference between the mean response latency for AUD and that for CAT was significant. Overall, the mean response latency for CAT was longer was longer than that for AUD. The accuracy rate between the two tasks was also significantly different. However, the longer latency for CAT did not lead to higher accuracy, compared with AUD. In terms of efficiency, the participants tended to perform better for AUD than for CAT. Therefore, we have come to a tentative conclusion that discrimination that involved categorical knowledge was more demanding. As a result, the participants responded more slowly, less correctly, and thus less efficiently. For AUD, where no categorical knowledge was required, the task could be achieved merely through the acoustic information of the pairwise tone stimuli.

Categorical perception of auditory signals is associated with listeners' experience with those sounds, as previous studies have observed (e.g., Francis, Ciocca, & Ng, 2003; Xi, Zhang, Shu, Zhang, & Li, 2010; Zheng et al., 2012), although lexical tone has been observed to be perceived categorically in tone languages such as Cantonese and Mandarin Chinese. Non-speech auditory signals cannot be perceived categorically (Liberman, 1982) Accordingly, it might be that the up-down contour (i.e., the invented tone contour, Tone 5 in the experiment) did not exist in the tone repertoire of the native Mandarin Chinese listeners' mental lexicon, so the native Chinese listeners needed to record the acoustic waveform into a more abstract category before a category discrimination could be done. The additional cognitive process might be one of the intervening factors that contributed to longer latency for the category discrimination when the participants were forced to make a judgment.

3.2 *Speech-likeness of Sound*

3.2.1 *Performance between SW and LX* The *Sound* types, the sine-wave tone (SW) and the lexical tone (LX) did not make a significant difference in the mean repose latency of the discrimination tasks, $F(1, 1310) = 0.54$, $p > .05$. The mean discrimination latency at SW (M = 740 ms, SD = 376 ms) was not significantly different from that at LX (M = 752 ms, SD = 384 ms).

Despite no difference in the mean response latencies across the *Sound* types, the overall accuracy of the discrimination tasks was better at SW (accuracy = 81%) than at LX (accuracy = 74%). In terms of efficiency, that is, when accuracy is viewed in relation to time expenditure, overall discrimination was performed more efficiently at SW (*eff.* = 0.92) than at LX (*eff.* = 0.84).

Nevertheless, there was a significant three-way interaction existing between the independent variables, $F(1, 1310) = 4.19$, $p < .05$. The effects of the *Sound* types depended on the interaction *Context x Task,* as we see in the following analysis between and within levels of *Sound* in *Context x Task.*

We speculate that the pair-wise tone stimuli might be easier to discriminate in the mode of sine-wave tone than in the mode of lexical tone because of not being speech-like. The sine-wave tone stimuli were not speech-like in that they were rid of vocalic information (i.e., formats) except for the acoustic information of pitch. Very likely, it is this characteristic that led to better accuracy, and thus better efficiency at SW (*eff.* = 0.92) than at LX (*eff.* = 0.84), which was rich in lexical-semantic information in addition to acoustic and phonetic information. However, the superiority of efficiency of SW over LX did not prevail. On some conditions, the participants discriminated more efficiently at SW as they did at LX. On other conditions, the participants discriminated at SW as efficiently as at LX.

3.2.2 *Efficiency in Task x Context within and between SW and LX* Of the four conditions at SW, the efficiency in AUD *x* NF *x* SW (*eff.* = 0.27) was the best. In contrast, of the four conditions at LX, AUD *x* NF *x* LX (*eff.* = 0.24) had the best efficiency. Nevertheless, the superiority of SW over LX in discrimination efficiency does not hold in every condition. For example, in AUD *x* FF, the participants performed as efficiently at SW (*eff.* = 0.23) as at LX (*eff.* = 0.23). This reveals that the sine-wave tone stimuli and the lexical tone stimuli might be perceived similarly for AUD at FF contexts.

Further post-comparison analyses reveal that discrimination performance at SW and at LX differed in effect depending on the interaction of *Task x Context.* The effects of the interaction of *Task x Context* within and between levels of *Sound* were examined at length. In the following section, post-comparison across conditions reveals how discrimination was performed differently at SW in comparison to that at LX.

3.2.3 *Implications Task x Context: SW vs. LX* In AUD *x* FF, the *Sound* types made no significant difference in the response latency. The difference of the mean response latency (after rounding-off) between

SW (771 ms) and LX (771 ms) was nearly nought. Also, in terms of efficiency, for AUD, the discrimination at SW was performed as efficiently as that at LX. The discrimination efficiency in FF *x* AUD *x* SW (*eff.* = 0.23) equaled that in FF *x* AUD *x* LX (*eff.* = 0.23). That being the case, in AUD *x* FF, the native Mandarin Chinese listeners performed the discrimination in the mode of SW and in the mode of LX equally efficiently when the contours were familiar to them. Previous studies have shown that sine wave speech could be taken as speech or non-speech, depending on the previous language experience of the listeners (Möttönen et al., 2006; Remez et al., 1981). The result implies that at FF contexts, pair-wise stimuli of sine-wave tones were no less efficiently discriminated between than pair-wise lexical tones. The familiarity with the tone contours might play an effective role in the auditory discrimination. Even though the auditory discrimination did not require categorical knowledge of tone, the familiarity with the tone contours might have implicitly activated the categorical knowledge about them.

In contrast, a divergence arose in CAT *x* FF. In CAT *x* FF, discrimination was performed more efficiently at SW than at LX. It is probable that lexical tones were far more complex than sine-wave tones in that lexical tones not only involve acoustic information of the sound signal, but also lexical, semantic, or affective connotation, even articulatory preparation. The sine-wave tone stimuli might not activate so much information compared with the lexical tone stimuli, so discriminating between the pair-wise sine-wave tone stimuli was not as demanding as discriminating between pair-wise lexical tone stimuli.

However, in CAT *x* NF, *Sound* types made no difference in the discrimination efficiency at SW (*eff.* = 0.22) and at LX (*eff.* = 0.22). The discrimination efficiency in NF *x* SW *x* CAT (*eff.* = 0.22) equaled that in NF *x* LX *x* CAT (*eff.* = 0.22). According to Liberman (1982), non-speech sounds were not processed categorically, and therefore, lexical tones carrying the novel contour might have been perceived as non-speech signals, since the invented tone contour (i.e., the up-down sweep) was not in the participants' tone repertoire.

Among the four conditions of the interaction *Task x Context* with *Sound,* as demonstrated in Figure 6, the participants discriminated SW as efficiently as they discriminated LX on Condition (a) and Condition (b), while on Condition (c) and Condition (d), the participants performed more efficiently at SW than at LX. The participants' discrimination performance for each condition is discussed as follows.

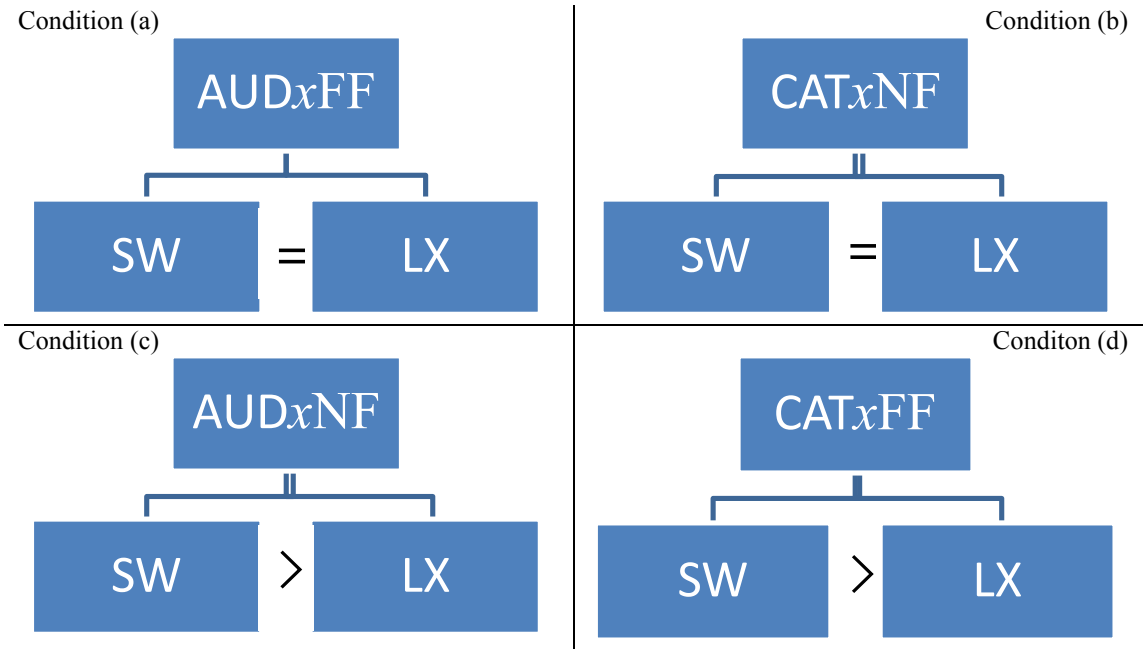

Figure 6. The interaction of *Task x Context* at each level of *Sound:* SW vs. LX

On Condition (a), AUD *x* FF, where the pair-wise stimuli comprised two familiar tone contours, categorical knowledge might have been implicitly activated, and consequently, SW, just like LX, was implicitly perceived categorically by the participants. On Condition (b), CAT *x* NF, the pair-wise stimuli, comprising one exemplar of the invented tone contour and one exemplar of familiar tone contours, might not have been perceived categorically because the invented tone contour was not existent in the participants' tone repertoire. Lexical tones with the invented tone contour might very likely be perceived as non-speech signals.

Nevertheless, SW and LX had more effects on the discrimination tasks on Condition (c) and on Condition (d). SW still had an advantage over LX. Overall, it was easier to discriminate SW than LX probably because of the decreased complexity of sine-wave tones. The complexity of lexical tones tax more on the auditory processing capability of the brain, compared to sine-wave tones.

To wrap up, overall discrimination efficiency was better at SW than at LX. However, a significant three-way interaction *Task x Context x Sound* existed. In AUD *x* FF, when the participants were familiar with the tone contours, the participants performed AUD as efficiently at SW (*eff.* = 0.23) as they did at LX (*eff.* = 0.23), whereas in CAT *x* FF, the participants performed CAT more efficiently at SW (*eff.* = 0.21) than at LX (*eff.* = 0.15). In contrast, in CAT *x* NF, when the participants were not familiar with the invented tone contours, the participants performed CAT as efficiently at SW as they did at LX, while in AUD *x* NF the participants performed CAT more efficiently at SW (*eff.* = 0.27) than at LX (*eff.* = 0.24).

3.3 *Novelty of Context*

3.3.1 *Performance between FF and NF* *Context,* indicating the construct of the pair-wise tone stimuli, comprised two levels: familiar-familiar contour contexts (FF) and novel-familiar contour contexts (NF). The mean discrimination latency at FF (M = 781 ms, SD = 383 ms) was significantly longer than that at NF (M = 712 ms, SD = 373 ms), $F(1, 1310) = 11.66$, $p < .05$. The overall accuracy of the discrimination tasks at NF (79%) was higher than that at FF (76%). In terms of efficiency, discrimination was better performed at NF (*eff.* = 0.95) than at FF (*eff.* = 082).

Discrimination efficiency at NF was the better. We have come to the inference that the invented tone contour in the pairwise tone stimuli might be an intervening factor that led to better discrimination performance. The result implies that the participants were more sensitive to the invented tone token at NF than at FF. The novelty of the invented tone might increase perceptual sensitivity to the difference between the pair-wise tone stimuli. In turn, the increased auditory sensitivity might have positively affected the discrimination of the pairwise tone tokens. Very likely due to the increased sensitivity, the participants performed discrimination more efficiently at NF than at FF.

3.3.2 *Efficiency in Sound x Task within and between FF and NF* Of the four conditions at FF, discrimination efficiency in SW *x* AUD *x* FF (*eff.* = 0.23) and that in LX *x* AUD *x* FF (*eff.* = 0.23) was equal. That being the case, at FF, being speech-like or not made no difference for AUD. Therefore, at FF, SW and LX for AUD might be treated similarly by the native Mandarin listeners. Of the four conditions at NF, discrimination efficiency in SW *x* AUD *x* NF was the best. At NF, we observed that *Sound* types of the tone stimuli did not make a difference in discrimination efficiency between SW *x* CAT *x* NF (*eff.* = 0.22) and LX *x* CAT *x* NF (*eff.* = 0.22). At NF, SW and LX for CAT might be processed similarly. Comparisons across levels of *Contour* in *Sound x Task* show that discrimination task tended to be conducted more efficiently at NF than at FF. Novelty of the invented tone contour might have exerted a decisive influence.

3.3.3 *Implications--Sound x Task: FF vs. NF* The difference between the mean response latency at FF and that at NF was significant. The mean response latency at FF was longer than that at NF. The participants also achieved more accurate discriminations at NF than at FF. When we view the accuracy rate in relation to time expenditure, that is, in terms of efficiency, the participants performed better at NF than at FF. The native Chinese listeners' familiarity with the tone contours at FF seemed not to have given them an advantage over the novelty of the invented tone contour at NF during the discrimination tasks. The invented tone contour of tone was not existent in the participants' tone repository. Such an invented tone contour of tone did not exist since they had been no exposure to the invented tone contour in their ambient language environment.

This results indicate that discriminating the pairwise tone stimuli at FF was more demanding than at NF. We speculate that the contrast between "novel" and "familiar" in the pairwise tone stimuli at NF was more prominent than that at FF to the Mandarin Chinese participants, so discrimination was more efficiently

performed at NF. Novelty of the invented tone contour might have made it easier for the participants to distinguish the novel contour from the familiar contour in the pairwise tone stimuli.

3.4 *Optimal auditory perception*

To account for the superiority of one condition over the other, we propose that auditory signals are perceived optimally when the cost of processing is the lowest. We refer to this proposition as *Optimal Auditory Perception* (OAP). The proposition prescribes that acoustic signals are perceived most efficiently when the cost of processing is the lowest. The key concept of optimality is that speech perception is a balance between time and accuracy in the current experiment, where the participants had to perform a tone discrimination task as fast and accurately as possible. The trick of performing the discrimination task is to strike a balance between latency and accuracy in such a way that discrimination efficiency was maximized. We define the cost of processing in terms of constrain violations, in the terminology of *Optimality Theory* (McCarthy, 2001). The more constrain violations, the more costly the processing. In the section to follow, we first devised constrains based on the results of the ANOVA, and then applied the proposed constrains to account for the optimal performance of discrimination in certain conditions on the basis of efficiency.

Constrains are the specified situations that affect the participants' capability to maximize the efficiency of speech perception. The results of ANOVA suggest that complexity of *Task*, and speech-likeness of *Sound*, novelty of *Context* are factors that influence the efficiency of discriminating between the pairwise tone stimuli. According to the results of ANOVA, we propose three constrains, **Categorical*, **Speech-like*, and *Novel* to account for the optimal performance of discriminating the pairwise tone stimuli in specific conditions. **Categorical* is associated with the nature of *Task*, **Speech-like* is related to the speech-likeness of *Sound*, and Novel is about the novelty of *Context,* because it involved an invented tone. These three proposed constrains are defined as follows:

***Categorical:** For discrimination to be performed efficiently, categorical knowledge must not be involved.
***Speech-like:** For discrimination to be performed efficiently, tone stimuli must not be speech-like.
Novel : For discrimination to be performed efficiently, the context of the pair-wise stimuli must be novel.

OAP analysis with the proposed constrains and their ranking provides an account for the optimal winner and the ultimate loser within and between levels of the independent variables. **Categorical* plays the most influential role in deciding the winner. Any candidate violating **Categorical* is doomed. Candidates that violate *Novel* still stand a chance to be the winner. Novelty of tone contour played a role in the category discrimination by the native Mandarin Chinese listeners. However, **Speech-like* did not always have a crucial effect on the discrimination efficiency probably because sine-wave tones might sometimes be perceived as lexical tones by the native Mandarin Chinese listeners. In compliance with **Categorical, *Speech-like*, and *Novel*, Candidate c, NF *x* SW *x* AUD, stands out as the optimal candidate--the winner because it abides by all of the three constrains. And because of violating all of the three constrains, Candidate f, FF *x* LX *x* CAT, ends up as the ultimate loser. The final ranking argument goes as follows: **Categorical* dominates *Novel*, which dominates **Speech-like*, as illustrated in the summary tableau.

*Summary Tableau: *Categorical>> Novel >>*Speech-like*

Candidates		Efficiency	*Categorical	Novel	*Speech-like
☞	c NF *x* SW *x* AUD	**0.27**			
	d NF *x* LX *x* AUD	0.24			*
	a FF *x* SW *x* AUD	0.23		*	
	b FF *x* LX *x* AUD	0.23		*	
	g NF *x* SW *x* CAT	0.22	*		
	h NF *x* LX *x* CAT	0.22	*		
	e FF *x* SW *x* CAT	0.21	*	*	
Ω	f FF *x* LX *x* CAT	0.15	*	*	*

Note. ☞indicates the optimal performer; Ω indicates the worst performer.

4 Concluding Remarks

The results of AVOVA show that *Task* and *Context* had significant effects on the response latencies. The same tendency still holds in terms of efficiency. The *Task* types had a decisive influence on the discrimination performance, with the auditory discrimination more efficiently performed than the category discrimination. The *Context* types also had a significant influence on the discrimination efficiency, with discrimination better performed when the pair-wise tone stimuli comprised one novel tone contour and one familiar tone contour, than they were made up of two familiar contours.

When it comes to *Sound*, the overall discrimination tended to be more efficient in the mode of sine-wave tone than in the mode of lexical tone. Nevertheless, being speech-like or not did not always have a crucial effect on the discrimination efficiency. Whether the *Sound* types made a significant difference was dependent on the interaction of *Task x Context,* seeing that there existed a significant three-way interaction between the independent variables.

The present study observed that sine-wave tones might be perceived as lexical tones to the native Chinese listeners, but only on Condition (a) that the contexts were familiar to the listeners for the auditory discrimination task, where no categorical knowledge was involved, and on Condition (b) that the contexts were novel to the listeners for the category discrimination task, where categorical knowledge was required. Only on these two conditions would the pair-wise sine-wave tone stimuli be discriminated as efficiently as the pair-wise lexical tone stimuli. The results are also accounted for by the proposed *Optimal Auditory Perception.*

Nevertheless, the discrimination tasks of the current experiment on tone perception might also involve conscious comparison, categorization, memory work, or other cognitive functions. Further studies are needed to tear apart these cognitive functions, and to reveal neural substrates underlying the perception of lexical tone. More evidence, neural evidence in particular, is necessary to verify the psychological reality of these behavioral results.

The mechanism of *Optimal Auditory Perception* (OAP) is borrowed mainly from *Optimality Theory* (OT), where constrains and the cost of constrain violation play important role in the output of speech production. Seldom has OT theory been applied to explaining speech perception in the past. The application of OAP to explaining speech perception is a preliminary attempt to account for the factors that affect auditory perception of human speech system. Although the proposed constrains are not universal, as OT requests, the practice of OAP provides a feasible account of tone discrimination by normal listeners of Mandarin Chinese. However, more elaboration is needed before the application of OAP can become full-fledged.

References

Boersma, P. & Weenink, D. (2013). *Praat*: doing phonetics by computer [Computer program]. Version 5.3.51. Retrieved from http://www.praat.org/

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

Frost, J. A., Binder, J. R., Springer, J. A., Hammeke, T. A., Bellgowan, P. S., Rao, S. M., & Cox, R. W. (1999). Language processing is strongly left lateralized in both sexes. Evidence from functional MRI. *Brain, 122 (2)*, 199-208.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Satthamnuwong, N., & Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language, 84*, 318-336.

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

Klein, D., Zatorre, R. J. Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage, 13,* 646-653.

Liberman, A. M. (1982). On finding that speech is special. *American Psychology, 37,*148-167.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neutral substrates of phonemic perception. Cerebral Cortex, 15, 1621-1631.

McCarthy, J. (2001). A Thematic Guide to Optimality Theory. Cambridge: Cambridge University Press.

Möttönen, R, Calvert G. A., Jääskeläinen I. P., Matthews, P. M., Thesen, T., Tuomainen, J., & Sams, M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage, 30 (2)*, 563-569.

Qiu, Gui-Su. (2012). How to Pronounce Mandarin Chinese--Mandarin Audio Pronunciation Guide. Retrieved Nov. 11, 2012, from http://mandarin.about.com/od/pronunciation/a/How-To-Pronounce-Mandarin-Chinese.htm

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947-949.

SPSS Inc. Released 2009. PASW Statistics for Windows, Version 18.0. Chicago: SPSS Inc.

Tervaniemi, M. & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews, 43,* 231-246.

Whalen, D. & Liberman, A. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237*, 169-171.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience, 170, 223-231*

Zatorre, R. (2003). Hemispheric asymmetries in the processing of tonal stimuli. In K. Hugdahl & R. J. Davision (Eds.), *The Asymmetrical Brain* (pp. 411-440). Cambridge, MA: MIT Press.

Zheng, H-Y., Minett, J. W., Peng, G. & W. S-Y. Wang. (2012). The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Language and Cognitive Processes, 27 (2),* 184-209.