Crossmodal correspondences between linguistic tones and complex 3-D shapes and their low-level sensory properties (Pitch & spatial frequency)

Shang, Nan

2017

Shang, N. (2017). Crossmodal correspondences between linguistic tones and complex 3-D shapes and their low-level sensory properties (Pitch & spatial frequency). Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/72028>

<https://doi.org/10.32657/10356/72028>

**CROSSMODAL CORRESPONDENCES BETWEEN
LINGUISTIC TONES AND COMPLEX 3-D SHAPES
AND THEIR LOW-LEVEL SENSORY PROPERTIES
(PITCH & SPATIAL FREQUENCY)**

SHANG NAN

SHANG NAN

School of Social Sciences

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

2017

Acknowledgements

My special thanks to Dr. Suzy Styles, my supervisor, who introduced this area to me and guided me down this road. Without her guidance and patience, this work would not have been possible. I would also like to thank Dr. Cho Sook Whan and Dr. Kwon Nayoung for introducing me to this program. Thanks also go to Dr. Randy J. Lapolla who has been so kind and willing to help at the hardest times during this journey. Thanks are also due to Dr. Maus, Gerrit for his help with the code for spatial frequency and Dr. James German for his help with the acoustic stimuli recording. I would also like to give thanks to the faculty members and the staff of the Psychology Division at Nanyang Technological University for their guidance and help over the years. Many thanks are also due to all the participants for their participation in my experiments. Great thanks also go to the members of my thesis committee for their time and valuable comments. I also thank my graduate colleagues as well as all the friends I made in Singapore for accompanying me on this adventure. Finally, thanks and love to my parents and J 小 Q who always support and believe in me. This has been a fun and tough ride, and it's great to share it with all of you.

Table of Contents

List of Tables	vi
List of Figures	viii
Abstract	x
Chapter 1: Introduction	1
1.1 Phonetic symbolism in natural languages	1
1.2 Sound symbolism when using Nonwords	5
1.3 The Bouba-kiki effect in toddlers and infancy	10
1.4 Sound-symbolic words in natural language—Ideophones	14
1.5 Which sounds go with which shapes?	15
1.6 Does the bouba-kiki effect have a synaesthetic origin?	16
1.7 Relationships between articulation mechanism and physical properties of objects	19
1.8 ERP and fMRI Evidence for sound symbolism	21
1.9 Developmental disorders & poor sound-symbolic performance	24
1.10 Systematicity outside language: pitch effects	26
1.11 Linguistic experience tunes speech perception	36
1.12 Pitch in language	39
1.13 Sound symbolism in the Chinese language	43
1.14 The current study—Mandarin-tone-related Sound Symbolism	46
1.15. Analytical Methodology	49
Chapter 2: Two-Alternative Forced Choice (2AFC) Tasks of Shapes and Vowels with Mandarin Chinese Tones	57
2.1 Introduction	57
2.2 Study 1A: Two shapes with one sound	58
Participants	58
Stimuli	61
Auditory stimulus validation	62
Experimental platform	64
Procedure	64
Predictions	64
Results	65
Discussion	79
2.3 Study 1B: Two sounds with one shape	81
Participants	81
Stimuli	81

Procedure	82
Results.....	82
General Discussion	93
Chapter 3: IAT of Shapes and Vowels with Tones.....	95
3.1 Introduction.....	95
3.2 Study 2A: IAT of Shapes and Vowels in Tone 1 & 4 – Vowel Attended	96
Participants.....	96
Apparatus	97
Stimuli.....	97
Procedure	97
Data handling and analysis	102
Results.....	103
Discussion	109
3.3 Study 2B: IAT of Shapes and Vowels in Tone 1 & 4 – Tone Attended	110
Results.....	111
Discussion	119
3.4 Study 2C: IAT of Shapes and Vowels in Tone 1 & 3 – Vowel Attended.....	120
Results.....	122
Discussion	127
3.5 Study 2D: IAT of Shapes and Vowels in Tone 1 & 3 – Tone Attended.....	128
Results.....	130
Discussion	136
3.6. Re-analysis of Study 2 using the D measure.....	139
3.7 General Discussion	149
Chapter 4: IAT of Shapes and Sine Waves with Tones	151
4.1 Introduction.....	151
4.2 Study 3A: IAT of Shapes and Sine Waves in Tone 1 & 4.....	152
Participants.....	152
Stimuli.....	152
Procedure	153
Results.....	155
Discussion	159
4.3 Study 3B: IAT of Shapes and Sine Waves in Tone 1 & 3	160
Participants.....	161
Stimuli.....	161

Procedure	161
Results.....	162
Discussion	166
4.4 Study 3C: IAT of Shapes and Sine Waves with intensity in Tone 1 & 4	167
Results.....	168
Discussion	173
4.5 A comparison between Study 2B & Study 3C.....	173
4.6 Study 3D: IAT of Shapes and Sine Waves with intensity in Tone 1 & 3	179
Results.....	180
Discussion	184
4.7 A comparison between Study 2D & Study 3D	186
4.8 General Discussion	192
Chapter 5: IAT of Visual Spatial Frequency and Sounds with Mandarin Chinese Tones	194
5.1 Introduction.....	194
5.2 Study 4A: IAT of Visual Spatial Frequency and Vowels in Tone 1 & 4 – Vowel Attended	195
Participants.....	195
Stimuli.....	196
Procedure	197
Results.....	199
Discussion	205
5.3 Study 4B: IAT of Visual Spatial Frequency and Vowels in Tone 1 & 4 – Tone Attended	205
Participants.....	206
Procedure	206
Results.....	207
Discussion	213
5.4 Study 4C: IAT of Visual Spatial Frequency and Sine Waves with intensity in Tone 1 & 4	215
Participants.....	215
Procedure	215
Results.....	216
Discussion	222
5.5 A comparison between Study 4B & Study 4C.....	223
5.6 General Discussion	229
Chapter 6: Conclusion.....	231
6.1 Language specificity and language acquisition.....	232
6.2 Cross-modal correspondence between vowels, visual angularity and Mandarin tones	233

2AFC tasks (Study 1A & 1B) using lexical tones.....	233
IAT tasks using lexical tones (Study 2A, 2B, 2C, 2D)	234
Non-lexical-tone tasks (Study 3A, 3B, 3C, 3D).....	237
6.3 Mappings between visual spatial frequency and Mandarin tones (Study 4A, 4B, 4C).....	239
6.4 Implications.....	240
6.5 Limitations and future directions	243
6.6 General Conclusions	245
References.....	247
Appendices.....	258

List of Tables

Table 1.1. Average F0 values (in Hz) of [u] and [i] produced in the four Mandarin tones by a female native Mandarin speaker	48
Table 2.1. Self-reported language levels.....	60
Table 2.2. Rating of typicality of the auditory stimuli	63
Table 2.3. Three groups' Binominal Test results of the pointy stimulus compare to the chance level	67
Table 2.4. GLMM Fixed Effects (Study 1A, overall).....	72
Table 2.5. Descriptive statistics of the pitch of the eight sounds used in this study produced by a female native Mandarin speaker (Hz).....	73
Table 2.6. GLMM Fixed Effects (Study 1A, /i/).....	74
Table 2.7. GLMM Fixed Effects (Study 1A, /u/).....	76
Table 2.8. GLMM Fixed Effects (/u/, CE : ENG)	77
Table 2.9. Pairwise comparisons between tones.....	77
Table 2.10. GLMM Fixed Effects (/u/, C: ENG).....	78
Table 2.11 GLMM Fixed Effects (/u/, C: CE).....	78
Table 2.12. GLMM Fixed Effects (Study 1B, T1-T3, overall).....	84
Table 2.13. GLMM Fixed Effects (Study 1B, T1-T3, CE vs. E).....	85
Table 2.14. GLMM Fixed Effects (Study 1B, T1-T3, C vs. E)	86
Table 2.15. GLMM Fixed Effects (Study 1B, T1-T3, C vs. CE).....	87
Table 2.16. GLMM Fixed Effects (Study 1B, T1-T4, overall).....	88
Table 2.17. GLMM Fixed Effects (Study 1B, T1-T4, CE vs. E).....	89
Table 2.18. GLMM Fixed Effects (Study 1B, T1-T4, C vs. E)	90
Table 2.19. GLMM Fixed Effects (Study 1B, T1-T4, C vs. CE).....	90
Table 3.1. Different conditions of Vowel-angularity congruence and Tone-angularity congruence in Study 2A.....	104
Table 3.2. ANOVA results of iRT (Chinese-dominant bilinguals).....	106
Table 3.3. ANOVA results of iRT (Chinese-English balanced bilinguals)	107
Table 3.4. ANOVA results of iRT (non-Chinese English-speaking participants)	108
Table 3.5. ANOVA results of iRT (overall Study 2A)	109
Table 3.6. ANOVA results of iRT (Chinese-dominant bilinguals).....	113
Table 3.7. ANOVA results of iRT (Chinese-English balanced bilinguals)	115
Table 3.8. ANOVA results of iRT (non-Chinese English-speaking participants)	116
Table 3.9. ANOVA results of iRT (overall Study 2B)	118
Table 3.10. ANOVA results of iRT (Chinese dominant bilinguals).....	123
Table 3.11. ANOVA results of iRT (Chinese-English balanced bilinguals)	124
Table 3.12. ANOVA results of iRT (non-Chinese English-speaking participants)	126
Table 3.13. ANOVA results of iRT (overall Study 2C)	127
Table 3.14. ANOVA results of iRT (Chinese dominant bilinguals).....	131
Table 3.15. ANOVA results of iRT (Chinese-English balanced bilinguals)	132
Table 3.16. ANOVA results of iRT (non-Chinese English-speaking participants)	133
Table 3.17. ANOVA results of iRT (overall Study 2D)	135
Table 4.1. ANOVA results of iRT (Chinese-English balanced bilinguals)	157
Table 4.2. ANOVA results of iRT (non-Chinese English-speaking participants)	158

Table 4.3. ANOVA results of iRT (overall Study 3A)	159
Table 4.4. ANOVA results of iRT (Chinese-English balanced bilinguals)	164
Table 4.5. ANOVA results of iRT (non-Chinese English-speaking participants)	165
Table 4.6. ANOVA results of iRT (overall Study 3B)	166
Table 4.7. ANOVA results of iRT (Chinese dominant bilinguals).....	169
Table 4.8. ANOVA results of iRT (Chinese-English balanced bilinguals)	170
Table 4.9. ANOVA results of iRT (non-Chinese English-speaking participants)	171
Table 4.10. ANOVA results of iRT (overall Study 3C)	172
Table 4.11. ANOVA results of interactions between Tone Congruence, Speech type and Language group (Study 2B vs. 3C).....	178
Table 4.12. ANOVA results of iRT (Chinese dominant bilinguals).....	182
Table 4.13. ANOVA results of iRT (Chinese-English balanced bilinguals)	183
Table 4.14. ANOVA results of iRT (non-Chinese English-speaking participants)	184
Table 4.15. ANOVA results of iRT (overall Study 3D)	185
Table 4.16. ANOVA results of interactions between Tone Congruence, Speech type and Language group (Study 2D vs. 3D)	191
Table 5.1. ANOVA results of iRT (Chinese dominant bilinguals).....	201
Table 5.2. ANOVA results of iRT (Chinese-English balanced bilinguals)	201
Table 5.3. ANOVA results of iRT (non-Chinese English-speaking participants)	202
Table 5.4. ANOVA results of iRT (overall Study 4A)	204
Table 5.5. ANOVA results of iRT (Chinese dominant bilinguals).....	209
Table 5.6. ANOVA results of iRT (Chinese-English balanced bilinguals)	210
Table 5.7. ANOVA results of iRT (non-Chinese English-speaking participants)	211
Table 5.8. ANOVA results of iRT (overall Study 4B)	212
Table 5.9. ANOVA results of iRT (Chinese dominant bilinguals).....	218
Table 5.10. ANOVA results of iRT (Chinese-English balanced bilinguals)	219
Table 5.11. ANOVA results of iRT (non-Chinese English-speaking participants)	220
Table 5.12. ANOVA results of iRT (overall Study 4C)	222
Table 5.13. ANOVA results of interactions between Tone Congruence, Speech type and Language group (Study 4B vs. 4C).....	228

List of Figures

Figure 1.1. Time by frequency representations of [u] uttered in Mandarin Chinese tones by a female native Mandarin speaker (Tone 1: black line; Tone 2: yellow line; Tone 3: blue line; Tone 4: red line) (see p.44 & p52 in the following Chapter for details).	40
Figure 2.1. Visual stimuli of Study 1A	61
Figure 2.2. Pitch tracks of the eight auditory stimuli used in Study 1A ([i] stimuli: red lines; [u] stimuli: black lines). The figure was produced using PRAAT.	63
Figure 2.3. Angularity selection (%) for the three groups for Study 1A.	66
Figure 2.4. Estimated Means for significant effects (0=pointy) (Study 1A: overall)	71
Figure 2.5. Estimated Means for significant effects (0=pointy) (Study 1A: /i/)	74
Figure 2.6. Estimated Means for significant effects (0=pointy) (Study 1A: /u/)	75
Figure 2.7. Estimated Means for significant effects (0=T3) (Study 1B, T1-T3: overall)	84
Figure 2.8. Estimated Means for significant effects (0=T4) (Study 1B, T1-T4: overall)	88
Figure 3.1. Illustration of stimulus-response key assignment for practice trials.	98
Figure 3.2. The timeline for experimental trials in Study 2A	100
Figure 3.3. Examples of instructions used in the experimental blocks in Study 2A. Left: An example of the instruction for an angularity-vowel congruent block. Right: An example of the instruction for an angularity-vowel incongruent block.	102
Figure 3.4. Bar charts for different subject groups' back-transformed RTs in Study 2A (IAT of Shapes and Vowels in Tone 1 & 4: Vowel Attended) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)	105
Figure 3.5. Examples of instructions used in the experimental blocks in Study 2B for the two Chinese groups. Left: An example of the instruction for an angularity-tone congruent block. Right: An example of the instruction for an angularity-tone incongruent block.	111
Figure 3.6. Bar charts for different subject groups' back-transformed RTs in Study 2B (IAT of Shapes and Vowels in Tone 1 & 4: Tone Attended) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)	112
Figure 3.7. Tone-angularity congruent condition versus incongruent condition in Study 2B (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)	119
Figure 3.8. Bar charts for different subject groups' back-transformed RTs in Study 2C (IAT of Shapes and Vowels in Tone 1 & 3: Vowel Attended) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)	122
Figure 3.9. Examples of instructions used in the experimental blocks in Study 2D. Left: An example of the instruction for an angularity-tone congruent block. Right: An example of the instruction for an angularity-tone incongruent block.	129
Figure 3.10. Bar charts for different subject groups' back-transformed RTs in Study 2D (IAT of Shapes and Vowels in Tone 1 & 3: Tone Attended) (C: Chinese-dominant group; CE: Chinese-English balanced group; E: English group)	130
Figure 3.11. Tone-angularity congruent condition versus incongruent condition in Study 2D (C: Chinese-dominant group; CE: Chinese-English balanced group; E: English group)	136
Figure 4.1. The spectrogram of the original /u3/ used in the study (left) and the spectrogram of the sine wave of it with the measured pitch (right)	153

Figure 4.2. Examples of instructions used in the experimental blocks in Study 3A. Left: An example of the instruction for an angularity-tone congruent block. Right: An example of the instruction for an angularity-tone incongruent block.	154
Figure 4.3. Bar charts for different subject groups' back-transformed RTs in Study 3A (IAT of Shapes and Sine Waves in Tone 1 & 4) (CE: Chinese-English balanced group; ENG: English group).....	156
Figure 4.4. Examples of instructions used in the experimental blocks in Study 3B. Left: An example of the instruction for an angularity-tone congruent block. Right: An example of the instruction for an angularity-tone incongruent block.	162
Figure 4.5. Bar charts for different subject groups' back-transformed RTs in Study 3B (IAT of Shapes and Sine Waves in Tone 1 & 3) (CE: Chinese-English balanced group; ENG: English group)	163
Figure 4.6. Bar charts for different subject groups' back-transformed RTs in Study 3C (IAT of Shapes and Sine Waves with intensity in Tone 1 & 4) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group).....	168
Figure 4.7. Comparisons between lexical tones and sine waves (T1-T4) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)	175
Figure 4.8. Bar charts for different subject groups' back-transformed RTs in Study 3D (IAT of Shapes and Sine Waves with intensity in Tone 1 & 3) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group).....	181
Figure 4.9. Comparisons between lexical tones and sine waves (T1-T3) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)	188
Figure 5.1. Spatial frequency spectrum of the two visual stimuli used in our previous studies (Left: 0.0071 cycles/pixel; Right: 0.0177 cycles/pixel)	197
Figure 5.2. Visual stimuli used in Study 4 (Left: low spatial frequency; Right: high spatial frequency). 197	
Figure 5.3. Examples of instructions used in the experimental blocks in Study 4A. Left: An example of the instruction for a spatial frequency-vowel congruent block. Right: An example of the instruction for a spatial frequency-vowel incongruent block.	199
Figure 5.4. Bar charts for different subject groups' back-transformed RTs in Study 4A (IAT of Spatial Frequency and Vowels in Tone 1 & 4: Vowel Attended). (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group).....	200
Figure 5.5. Examples of instructions used in the experimental blocks in Study 4B. Left: An example of the instruction for a spatial frequency-tone congruent block. Right: An example of the instruction for a spatial frequency-tone incongruent block.	207
Figure 5.6. Bar charts for different subject groups' back-transformed RTs in Study 4B (IAT of Spatial Frequency and Vowels in Tone 1 & 4: Tone Attended). (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group).....	208
Figure 5.7. Tone-angularity congruent condition versus incongruent condition in Study 4B (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group).....	213
Figure 5.8. Examples of instructions used in the experimental blocks in Study 4C. Left: An example of the instruction for a spatial frequency-tone congruent block. Right: An example of the instruction for a spatial frequency-tone incongruent block.	216
Figure 5.9. Bar charts for different subject groups' back-transformed RTs in Study 4C (IAT of Spatial Frequency and Sine Waves with intensity in Tone 1 & 4). (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group).....	217
Figure 5.10. Comparisons between lexical tones and sine waves (T1-T4) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)	225

Abstract

This thesis investigates tone-related audiovisual cross-modal correspondences using lexical tones (i.e., Mandarin tones) for the first time. We used both 2AFC (two-alternative forced choice) tasks and IAT (Implicit Association Test) tasks involving three groups of participants (Chinese dominant bilinguals (C), Chinese-English balanced bilinguals (C/E), and English-speaking participants with no Chinese (E) (total N = 618 participants).

Chapter 1 outlines a summary of the literature on phonetic symbolism, ideophones, sound symbolism and synesthesia, neural evidence for sound symbolism and theories about the mechanisms of sound symbolism. Chapter 1 also reviews non-linguistic effects of pitch on cross-modal processing and how language experience tunes speech and tone perception. Throughout this literature, it is clear that lexical tones have rarely been examined in cross-modal correspondence studies.

In Chapter 2, using online 2AFC surveys, we replicated the ‘*u*-curvy, *i*-pointy’ vowel-angularity cross-modal correspondence and found that English-speaking participants (E) matched visual angularity with Mandarin tones in terms of pitch height (i.e., high-pointy, low-curvy). The Chinese dominant bilinguals (C) matched angularity with Mandarin tones in terms of pitch change (more change-pointy, less change-curvy). While Chinese-English balanced bilinguals (C/E) swung between the two strategies depending on task types.

To avoid possible effects of conscious strategy, in Chapter 3, we used IAT tasks to investigate crossmodal processing for lexical tone stimuli (i.e., /i1/, /i4/, /u1/, /u4/, /i3/, /u3/). We found significant group difference in this tone-shape crossmodal correspondence. The two Chinese groups consistently replicated the Pitch-Change Hypothesis (i.e., more change-pointy,

less change-curve) found in our 2AFC studies for the both tone contrasts (i.e., T1-T4, T1-T3) without significant group difference. On the other hand, the non-Chinese English-speaking group showed a significantly different tone-shape mapping pattern from that of the Chinese groups. The English speakers showed a ‘high-pointy, low-curve’ mapping pattern for both of the tone contrasts. This ‘high-pointy, low-curve’ crossmodal correspondence pattern was consistent with their preference observed in Study 1, where low tones go with curve shapes and high tones go with pointy shapes. These results were also consistent with the fact that pitch height is important in English speakers’ lexical tone perception (Gandour, 1983).

To see whether the spectral context of human voice makes difference in cross-modal correspondence between tones and shapes, in Chapter 4, we used sine waves with measured pitch (or both measured pitch and measured intensity) from the auditory stimuli used in the last chapter and IAT tasks. We found that all participant groups (i.e., C; C/E; E) matched visual angularity with tones in terms of pitch height without significant group difference. That is, pointy shapes were matched with high-pitched tones and curve shapes with low-pitched tones. In addition, the comparisons between Study 2 and Study 3 demonstrated that the difference in language-dependent effects between speech tones and non-speech tones was significant in the present sound symbolism tasks. That is, language specificity disappeared in cross-modal correspondence between visual angularity and non-lexical tones.

To see whether the spectral context of human voice makes a difference in cross-modal correspondence between spatial frequency and tones, in Chapter 5, we used both lexical and non-lexical tones to examine the relationships between tones and spatial frequency using the IAT paradigm. The Gabor patches used here were created using the modal spatial frequency information of the visual stimuli used in our previous studies. We found, for the first time, a

pattern of ‘*u*-low, *i*-high’ vowel-spatial frequency correspondence. When lexical tones were used, and tones were attended, both Chinese groups matched tones with spatial frequency in terms of pitch change (i.e., more pitch change with high spatial frequency, less pitch change with low spatial frequency) without significant group difference. However, the non-Chinese English-speaking group showed a significantly different mapping pattern from that of the two Chinese groups by pairing high spatial frequency with high pitch and low spatial frequency with low pitch. By contrast, when non-lexical tones were used, as seen in Chapter 4, all participants showed the same pitch-height-oriented cross-modal correspondence between tones and spatial frequency (i.e., high-high, low-low) without significant group difference, which was consistent with the previously observed cross-modal correspondence between pure tones and spatial frequency in English speakers (Evans and Treisman, 2010). Again, the difference in language-dependent effects between speech tones and non-speech tones was significant. Similar to what was observed in our angularity studies, language specificity again disappeared in tone-related audiovisual cross-modal correspondence when non-lexical tones were used.

In Chapter 6, we summarize that Chinese-speaking participants show ‘curvy-steady, pointy-dynamic’ mappings between shape and pitch only when linguistic stimuli are used, while the rest of time they show the ‘curvy-low, pointy-high’ pattern as non-Chinese participants do all the time. For mappings between spatial frequency and pitch, Chinese-speaking participants show ‘low-steady, high-dynamic’ pattern only when linguistic stimuli are used, while the rest of time they show the ‘low-low, high-high’ pattern as non-Chinese participants do all the time. For the first time, we examined cross-modal correspondences between shape and linguistic use of pitch. Theoretical and practical contributions of the present study are also highlighted in Chapter 6, along with some limitations of the present study and future research directions.

Chapter 1: Introduction

Most people have not thought about connections between senses, but their effects are surprisingly strong. Some of the connections have been explored, while others have not been explored. This chapter first introduces some debates on the existence of sound symbolism, which is followed by a review of empirical work using nonwords and sound-symbolic words. Investigations on the physical mechanisms and neural correlates underlying sound symbolism are also reviewed. Given that sound-shape correspondence is a core topic of this thesis, this chapter also reviews cross-modal mappings between sounds and shapes documented in the literature. Research on effects of pitch on sound symbolism has also been introduced, followed by an introduction to the current investigation of cross-modal correspondence between linguistic uses of pitch (Mandarin tones) and visual dimensions, for the first time.

1.1 Phonetic symbolism in natural languages

Saussure, one of the most influential modern linguists stated that the link between a linguistic sign and its referent is arbitrary (Saussure, 1916:1966). That is why different languages have different words for the same concept. Likewise, in ancient China, Xun Zi's argument that 'there are no correct names (名无固宜)', claims that there are no inherent correlations between a name and that which is named. A name is 'proper' as long as it is conventional. By

contrast, in the Cratylus dialogue, Socrates claims that a name specifies the nature of the named (Sedley, 2003). Jespersen (1922) also supports the idea that there are correspondences between certain types of sounds and certain kinds of meaning (i.e. sound/phonetic symbolism), such as size, luminance, distance, movement and feeling, by providing examples from the languages he was familiar with.

A great deal of word matching research has also demonstrated the existence of correspondences between sound and meaning across languages. Brown, Black and Horowitz (1955) asked subjects to guess meanings of foreign words from their sounds. They selected 21 pairs of English antonyms which were then translated into Chinese, Czech and Hindi by translators who were unaware of the purpose of the study. For each antonym pair, the subjects heard the foreign antonym pair and were then asked to match the two English words with their foreign language equivalents that they just heard. Most of the mean percentages correct for the three foreign languages were significantly better than chance ($p \leq .01$), suggesting that “there may be some features of phonetic symbolism which have a universal validity” (1955, p. 388).

Maltzman, Morrisett and Brooks (1956) conducted a series of similar experiments, asking English-speaking subjects to match words between English, Japanese and Croatian. In their experiments, they used words from one language as stimuli and words from another language as response pairs. One word of each response pair was the equivalent of the stimulus word. Subjects

were asked to pick one word from each response pair that seemed to best fit the corresponding stimulus word. In the experiments where English words were either stimuli or response pairs, the correct answers were chosen with a frequency significantly above chance expectancy, which was consistent with the findings of Brown, et al. (1955). However, they also found that when the participants' native language—English—was absent from the matching pairs, the matchings were not better than chance. The authors argued that the latter results did not undermine the general effect of sound symbolism, since the absence of the effect might be caused by the nature of the word matching tasks. However, the relationships between sounds and meanings are one-to-many. For example, the vowel [i] can be associated with small or sharp or bright meanings. When the native language is involved in the matching task, the respondents have a meaning, and then pick a best foreign word that matches that meaning. When the native language is not involved, without meanings as media, the relationships between the foreign words are many-to-many. Sounds can represent different meanings, the mapping between the foreign words could be harder.

Similarly, Brackbill and Little (1957) also claimed that their investigations have demonstrated the existence of sound symbolism. Using three methods of presentation (i.e., auditory only, auditory and visual and visual only), they asked their English-speaking subjects to pair English, Japanese, Chinese and Hebrew words. Contrary to Maltzman et al. (1956), their subjects paired Foreign-Foreign comparisons (e.g., Japanese-Chinese; Japanese-Hebrew) at

better than chance levels. They argue that similarity of word length, vowel, consonant types, etc. markedly influences word-meaning mappings across languages.

Atzet and Gerard (1965) asked Navajo-speaking subjects to choose one word from a verbally presented foreign (Chinese/Hindi) antonym pair for the Navajo word. The frequency of the correct choices was not significantly above chance. They argue that their non-significant results might be due to the absence of visual stimuli. For matching to be successful, visual stimuli are also very important, since they may provide additional cues aiding correct matching. Also, as seen in previous studies, it seems that some languages work better in foreign word pairings. Apparently, Navajo does not. Conversely, Klank et al. (1971) selected 16 antonym pairs and translated them into five foreign languages (Czech, Hindi, Chinese, Japanese and Tahitian). English-speaking subjects were asked to match foreign word pairs with English word pairs. To avoid phonetic cues, stimuli were presented visually. Subjects were able to guess the meanings of foreign antonym pairs above chance level. They argued such results lent support to the hypothesis of universal phonetic symbolism. An issue worth considering here is the role of selection bias, for example, one might get different results asking people to guess the meanings of the words “little” and “large” compared to “big” and “small”.

Sound-meaning correspondences also exist in the Chinese language. For example, 滴 is pronounced as dī and means ‘drop’. 溪 is pronounced as xī

and means ‘small stream, brook’. By contrast, 洋 is pronounced as yang2 and means ‘ocean’. These examples (extracted from Chan, 1996, p7) where small volumes of water are described with /i/ and large bodies of water with /a/, provide evidence to suggest that vowel-size correspondences exist in the Chinese language. Similar to the findings of investigations of sound-size correspondence using Western languages, /i/ matches with small things and /a/ matches with large things in the Chinese language as well.

Taken together, the existence of sound symbolism (phonetic symbolism) has been found in natural languages and proven experimentally.

1.2 Sound symbolism when using Nonwords

A great deal of research has been done on sound symbolism using nonwords. In Sapir’s seminal work (1929), subjects were provided with meaningless word pairs, each pair differing in vowels only, and asked to imagine which of the words would go with a larger or a smaller arbitrarily selected meaning. For example, the nonwords *mal* and *mil* were verbally presented and then subjects were asked to decide whether *mal* went with a large or a small table as compared with the word *mil*. There were 60 such nonword pairs and each pair differed only by one vowel. The results showed a high level of agreement on sound symbolism among subjects. For example, high (in terms

of tongue height), front (in terms of tongue backness) /i/ was consistently matched with small things and low, back /a/ was matched with large ones.

The other early influential phonetic symbolism experiment was conducted by Köhler (1929: 1947). In this study, Köhler showed subjects two figures, one blobby rounded and the other jagged star-like and asked the subjects which one was *maluma* and which one was *takete*. Most subjects chose *maluma* for the blobby line drawing and *takete* for the jagged one. According to Köhler's explanation, this is due to similarities between sounds and shapes. That is, the "roundness" of the sounds in *maluma* matches with the roundness of the rounded shape, etc.

Davis (1961) conducted a similar study by including English-speaking school children (11-14 years old) residing in England and Swahili-speaking school children (8-14 years old). The African children spoke Kitongwe (the Bantu language) at home and learned Swahili at school) living in an isolated peninsula in Africa. In this study, the researcher used the nonword *uloomu* instead of *maluma*, because *maluma* sounds like 'mother's brother' in Swahili. Children in the both groups consistently matched the curvy-shaped drawing with the nonsense name *uloomu* and the angular-shaped drawing with the nonsense name *takete*, suggesting the effect is cross-cultural and is evident from childhood.

Holland and Wertheimer (1964) investigated sound-shape mappings using Köhler's Maluma and Takete figures, as well as the nonsense words *Maluma*,

Takete, *Kelu*, and the letters U and K. According to some of their pilot studies, they argued that the letters U and K have “physiognomic properties” representing those of the entire words *Maluma* and *Takete*. The nonsense word *Kelu* was used, because it is composed of elements of both nonsense words. They asked adult subjects to rate how well each of the five verbal stimuli fitted the two line drawings. The results for verbal *Maluma* and *Takete* replicated Köhler’s study. The results for the letters U and K were highly similar to those of their corresponding full nonsense words, indicating phonetic symbolism can occur at single vowel or consonant level. Finally, as expected, *Kelu* fell in the middle. This might be because *Kelu* is half *Maluma* and half *Takete* or just due to chance responding.

Tarte and Barritt (1971) asked English-speaking adult subjects to decide which of two figures presented side by side went best with a verbally presented consonant-vowel-consonant (CVC)-type nonsense word. The figures they used were four ellipses and four triangles differing in size. 28 possible pairs of the eight figures were used as visual stimuli. For the CVC-type nonsense words, /w/, /d/ and /k/ were used as the initial letters, followed by /a/, /u/ or /i/, and all ended with /s/, ending up with nine nonsense words. Each sound was presented with a pair of geometric figures. The nonsense words were presented via audio tape and each slide of the visual stimulus pair was projected for the entire trial. The results showed that the subjects consistently assigned /a/ to large figures and /i/ to small ones. The subjects also paired ellipses with /u/, and triangles with /i/. /u/ did not seem strongly correlated with size, nor /a/ with shape. The

consonant effects were significant only for /k/-triangle matching and /w/-ellipse matching. This study replicated the effects found in Sapir (1929) and Köhler's (1929: 1947) studies with better controlled visual stimuli, but their auditory stimuli included the real word *was* and some stimuli similar to real words, e.g., *kis*, *dis*, which may give rise to semantic confounding. Also, the sample size here was extremely small ($N = 8$).

Employing Tarte and Barritt's (1971) stimuli and procedures, Tarte (1974) conducted the same investigation using Czech-speaking adults. Czech subjects showed the same phonetic symbolism pattern as their previous English-speaking subjects. That is, they consistently matched /a/ with larger figures, /i/ with small ones. They chose /u/ for ellipses, /i/ for triangles. As Czech and English are from the same language family, these findings supported Taylor and Taylor (1962)'s claim that speakers of related languages perform similarly in terms of phonetic symbolism. Moreover, Tarte suggests that oral cavity size, tongue position and distinctive features theory may account for the phenomenon.

Speaking of the generality of sound symbolism, a well-known experiment conducted by Ramachandran and Hubbard (2001) neatly replicated Köhler's (1929: 1947) findings using English-speaking American students and Tamil speakers in India. Their visual stimuli were a jagged figure and a rounded figure. The names they used were *bouba* and *kiki*. They found that overall 95-98% of the both groups assigned *bouba* to the rounded figure and *kiki* to the

jagged. As English (Indo-European Germanic language) and Tamil (Dravidian language) are from different language families, the similar phonetic symbolism patterns from speakers of these languages suggest the universality of the phenomenon between language families. Such results were consistent with Davis's findings (1961).

Similarly, when asked to match an angular shape and a rounded shape with the nonsense words *bouba* or *kiki*, the majority (28 out of 34; 82%) of the Himba participants matched the rounded shape with *bouba* and the angular shape with *kiki* (Bremner et al., 2013). The Himba of Northern Namibia are a remote population without a written language and with little exposure to Western culture. Such findings also show that the bouba-kiki effect is widespread across cultures.

To sum up, since the seminal work of Sapir (1929) and Köhler (1929: 1947), empirical research on sound symbolism using nonwords has been conducted cross-culturally with participants in different age groups. Consistent results have been observed. For example, in terms of sound-size mappings, /a/ is matched with largeness and /i/ with smallness; in terms of sound-shape correspondences, /u/ is matched with roundness and /i/ with angularity.

1.3 The Bouba-kiki effect in toddlers and infancy

The “bouba-kiki” effect has also been found in pre-reading toddlers. Maurer, Pathman, and Mondloch (2006) tested the bouba-kiki effect in 2.5-year-old children. They chose four pairs of contrasting shapes (one rounded, one angular). For each shape pair, they created a pair of nonsense words, in which one contained rounded vowels (e.g., bamu, pronounced ‘baa-moo’) and the other one contained non-rounded vowels (e.g., k[[^]]t[ej], pronounced ‘kuh-tay’). To maintain children’s interest, they tied the stimuli to a storyline. They used eight paired-preference questions (four validity check questions and four experimental questions). On each experimental trial, the experimenter told the child two names, e.g., Bamu and K[[^]]t[ej]. The experimenter also showed the child a jagged shape and a rounded shape and asked the child which shape he/she thought was *Bamu*. A group of non-synaesthetic adult participants were also recruited for comparison. The toddlers, like the control adults, matched rounded shapes to nonsense words containing rounded vowels and jagged shapes to nonwords containing non-rounded vowels. The authors argued that such results may suggest that sound-shape mappings can facilitate young children’s vocabulary acquisition and language learning. Their existing vocabulary may in turn influence the sound-shape mappings as well.

To investigate the role of *language experience* in the bouba-kiki effect, some researchers have conducted sound symbolism studies in preverbal infants. In a habituation-style looking time task for 4-month-old infants, Ozturk, Krehm,

and Vouloumanos (2013) created two large coloured shapes (one curvy, one angular). They used 'kiki' and 'bubu' as auditory stimuli. They created a 40-second 'kiki' string and a 40-second 'bubu' string, each of which contained 24 tokens. Based on the results of previous adult and toddler studies, they created four possible stimulus pairs in terms of sound-shape correspondence.

Congruent pairs: pairing the curvy shape with 'bubu' or the angular shape with 'kiki'; Incongruent pairs: pairing the curvy shape with 'kiki' or the angular shape with 'bubu'. Each pair of stimuli was presented to each infant until the infant looked away from the screen for 2 seconds or after a maximum string length of 40 seconds. Infants looked longer at incongruent pairs than at congruent ones. Using the same procedure, they further examined whether vowels (auditory stimuli: 'kiki' and 'kuku') or consonants (auditory stimuli: 'bubu' and 'kuku') alone can be matched to the shape of objects. However, no difference in looking time between congruent and incongruent pairs was found when auditory stimuli differed only in vowels, or only in consonants. Preverbal infants therefore showed sound-shape mapping discrimination (i.e., bouba-kiki effect) by looking differently between audiovisually congruent and incongruent pairs, and at this age these biases are only be observed for combinations of vowels and consonants.

Similarly, in a paired preference-style looking time study, Peña, Mehler, and Nespor (2011) synthesized monosyllables containing either high front vowels (i.e. /i/ or /e/) or low back vowels (i.e. /o/ or /a/) preceded by a consonant (/l/, /f/ or /d/). Each sound was paired with two shapes, one large and

one small, which were identical except for size. The shapes were outlines of geometric shapes (square, circle, oval or triangle) in one of four colors (red, yellow, green or blue). In each trial, infants were simultaneously presented with a pair of shapes differing only in size, and a sound repeated five times at one-second intervals. 4-month-old infants had more first-gazes and longer total looking time at large shapes than at small shapes when hearing syllables containing /o/ or /a/, and had more first-gazes and longer total looking time at small shapes than at large shapes when hearing syllables containing /i/ or /e/. In both this study and Ozturk et al. (2013), young infants showed discrimination ability for the congruent condition versus the incongruent condition. Ozturk et al (2013) attributed the difference in response patterns between the two studies, namely, longer looking time at incongruent pairings (Ozturk et al., 2013) and longer looking time at congruent trials (Peña, 2011), to difference in number of stimuli and mode of stimuli presentation between the two studies.

Some sound symbolism work in French adults (e.g., Fort, Martin, & Peperkamp, 2015) showed that French adults systematically mapped /m/ and /l/ onto rounded shapes and /k/ and /t/ onto spiky shapes. On the other hand, their mappings between vowels and shapes are less systematic (e.g., /o/ or /u/ with rounded shapes; /e/ or /i/ with spiky shapes). Fort, Weiss, Martin, and Peperkamp (2013) also investigated the presence of the bouba-kiki effect in French five-and six-month-old infants, and if so, whether vowels or consonants are more important in the bouba-kiki effect. They constructed 28 CVCV-type strings, half of which consisted of vowels and consonants that are all judged as

either rounded (e.g., /lumu/, /molo/, /lumo/, /mulo/) or spiky (e.g., /kiti/, /kete/, /keti/, /teki/). The other half consisted of either rounded vowels and spiky consonants (e.g., /kutu/, /took/, /kotu/, /kotu/) or rounded consonants and spiky vowels (e.g., /limi/, /mele/, /mile/, /lime/). They created a pair of outlined shapes (one rounded, one spiky). They then filled each shape with one of 7 colors (red, pink, yellow, green, brown, blue, purple) and mirrored each shape, ending up with 28 shapes. In each trial, infants were presented with a pair of shapes (one rounded, one spiky), followed after 500ms by the presentation of five repetitions of one auditory stimulus. Each trial lasted 10.5 seconds. Each infant was presented with 28 different trials divided into two blocks. They failed to find any bouba-kiki effect in any condition in terms of looking time. One possible reason for this result is the auditory stimuli contained both rounded vowels and spiky consonants in a single stimulus (e.g., *kotu*), or spiky vowels and rounded consonants in a single stimulus (e.g., *limi*), which might have reduced the likelihood of observing a symbolic effect. In spite of the difficulties in interpreting the null result for Fort et al. (2013), the general consensus is that multisensory correspondences have their origins in infant sensory processes and related attentional preferences (e.g., Mondloch & Maurer, 2004; Walker et al., 2010; Peña et al., 2011; Ozturk et al., 2013) which is consistent with the basic claim of Imai and Kita's (2014) sound symbolism bootstrapping hypothesis, that young infants are also sensitive to sound symbolism.

In conclusion, some sound-meaning mappings have been demonstrated in toddlers and preverbal infants as in adults, suggesting that such mappings precede language acquisition and may facilitate language development.

1.4 Sound-symbolic words in natural language—Ideophones

To avoid the potentially exaggerated perceptual differences between conditions caused by deliberate manipulation in made-up nonsense words, a great number of sound symbolism studies have been conducted on ideophones (or called mimetics). Ideophones were defined as either “lexicalized sound-imitative words” (Nuckolls, 1999) or “marked words which depict sensory imagery” (Dingemanse, 2012). Japanese is rich in ideophones. Imai, Kita, Nagumo, and Okada (2008) were interested in action-related Japanese sound-symbolic words. They created novel Japanese ideophones, and Japanese adults and English-speaking adults who have no knowledge of the Japanese language, as well as 3-year-old Japanese children successfully matched the sounds of the ideophones to their corresponding actions. Moreover, they found that 3-year-old Japanese children could even stably generalize the novel ideophones to new action events. Given that previous studies (Imai et al., 2006; Imai, Haryu, & Okada, 2005) reported that children at this age have difficulty generalizing other kinds of verbs, sound symbolism seems to play an important role in early verb learning. In a follow-up study, Kantartzis, Imai, and Kita (2011) conducted a verb generalization task with 3-year-old English-speaking children and found

that the English-speaking children learned novel sound-symbolic words better than sound-symbolic-mismatching words, suggesting that early language-independent sound symbolism facilitates early word learning. Moreover, in a word-learning task, Nygaard, Cook, and Namy (2009) found that English-speaking adults learned actual English equivalents and antonyms for Japanese words better than random pairs, suggesting non-arbitrary links between sound structure and meaning. Similar results were found by Lockwood, Dingemanse, and Hagoort (2016) when they taught and tested adult Dutch participants using Japanese ideophones and sound-symbolic-mismatching words. Such findings suggest that sound-symbolic links are universal and facilitate word learning in both young children and adults.

1.5 Which sounds go with which shapes?

Some researchers have taken their investigations of the bouba-kiki effect into which sounds are congruent with which shapes. As reviewed by Lockwood and Dingemanse (2015), both consonants and vowels influence participants' judgments of sound-shape mappings: Sonorants, voiced stops and back rounded vowels typically match with curvy shapes, voiceless stops and high-front unrounded vowels typically match with pointy shapes (c.f. Davis, 1961; D'Onofrio, 2013; Köhler 1929:1947; Maurer et al., 2006; Ozturk et al., 2013; Ramachandran & Hubbard, 2001; Nielsen & Rendall, 2011; 2013). For example, Nielsen and Rendall, (2011; 2013) found sonorant consonants (/m/,

/n/, or /l/) and rounded vowels (“oo”, “oh” or “ah”) were matched to curved images while voiceless plosive consonants (/t/, /k/ or /p/) and non-rounded vowels (“ee”, “ay” or “uh”) were matched to jagged images. Similarly, D’Onofrio (2013) found that non-words containing voiced consonants (/b/, /d/ or /g/), labial consonants (/b/ or /p/) and back and/or rounded vowels (/u/ or /a/) were matched with round shapes more than their respective counterparts (/t/, /k/; /i/, /e/). Moreover, Fort et al. (2015) also found French adults consistently matched /m/ and /l/ with rounded shapes while /k/ and /t/ with spiky shapes. By contrast, their mappings between /u/, /o/ and rounded shapes, as well as /i/, /e/ and spiky shapes were less systematic. Thus, Fort and her colleagues argued that consonants are more important than vowels in sound-shape correspondences. Taken together, such sound-shape mappings as ‘/u/-curvy’, ‘/o/-curvy’, ‘/m/-curvy’, ‘/i/-pointy’, ‘/e/-pointy’, ‘/k/-pointy’, ‘/t/-pointy’ have been consistently documented in previous studies.

1.6 Does the bouba-kiki effect have a synaesthetic origin?

Martino and Marks (2001) proposed that there may be strong and weak forms of synesthesia. As they put it, both types of synesthesia have evident cross-modal correspondences. A color-pain synesthete who sees colors in response to pain is an example of strong synesthesia. Weak synesthesia (or what is often called ‘cross-modal correspondences’) was observed in laboratory experiments, where most participants can match a stimulus from one sensory modality to a stimulus from another in a systematic manner. For example, most

people pair high-pitched sounds with smaller and more angular visual shapes (Marks, 1978). Given cross-modal correspondences in both type of synesthesia, Martino and Marks (2001) suggest that the neural substrates underlying both types of synesthesia may be in essence the same. However, at the end of their article, they admit that more neuroimaging evidence for whether the brain areas involved in the two type of synesthesia are similar are still in need. In line with this, Bankieris and Simner (2015) asked both grapheme-colour synaesthetes and non-synaesthete controls to listen to words from ten foreign languages and then guess their meanings from two alternatives. In the semantic domains of loud/quiet and big/small, both synaesthetes and non-synaesthetes showed above-chance-level accuracy, and synaesthetes performed significantly better than non-synaesthete controls, suggesting synaesthetes have greater sensitivity to the same type of sound symbolic correspondence than non-synaesthetes. Thus, they argue that synaesthesia is an exaggerated form of common cross-modal correspondences shared by regular people, rather than an essentially different phenomenon. Again, more neuroimaging evidence for neural difference and similarities between synaesthetes and non-synaesthetes is needed.

Moreover, given the findings from young infant studies, some researchers have proposed similar synesthetic-type explanations for the bouba-kiki effect. Maurer (1993) proposed that newborn infants are generally synaesthetic: She argued that “the newborn’s senses are not well differentiated but are instead intermingled in a synesthetic confusion” (Maurer, 1993: 111). Based on results of early infancy research (e.g., Hoffmann, 1978), Maurer (1993) suggested that

interactions between nearby areas in infants' primary sensory cortex are more prominent than in adults. Ramachandran and Hubbard (2001) agreed with Maurer's 'neonatal synesthesia model', arguing that the bouba-kiki effect in typical adults may be related to the sensory experiences in synaesthesia, – constituting a kind of sensory “cross-wiring” (e.g., perceiving numbers as coloured). They further speculated that synesthetic wiring may happen directly between sensory areas or mediated by the angular gyrus. Or, as Maurer and Mondloch (2005) put it, connections between cortical areas get pruned or inhibited later in development, but remaining cross-modal connections may be involved when normal children and adults match dimensions from different modalities (e.g., the bouba-kiki effect, or pitch-shape correspondence).

In an alternative line of reasoning, Deroy and Spence critically discuss the neonatal synaesthesia hypothesis. They claim that cross-modal correspondences observed in infants, toddlers and adults are apparent or superficial and there is insufficient evidence to support that innate synaesthetic confusion is the mechanism underlying them. They argue that the experience of synaesthesia is fundamentally different from that of cross-modal correspondence and therefore cross-modal correspondences found in infants, toddlers and adults cannot arise from a shared synaesthetic origin. By contrast, such accounts as statistical learning (Spence & Deroy, 2012) and equivalence learning (Deroy et al., 2014) can shed light on crossmodal correspondence (see Deroy & Spence, 2013, for a discussion of this idea). They also point out models of multisensory

development (Xu, Yu, Rowland, Stanford, & Stein, 2012) seem to be an appropriate explanation of the multisensory perception learning.

1.7 Relationships between articulation mechanism and physical properties of objects

Ramachandran and Hubbard (2001) further conjectured that lip and tongue movements during vocalization may be synaesthetically linked to objects and events they refer to. For example, words referring to something small often involve /i/ with narrowing of the vocal tract (e.g., words such as ‘little’, ‘teeny’). Nygaard et al. (2009) also argue that properties of sounds may relate to characteristics inherent in objects, events and etc. This cross-modal correspondence may be helpful for meaning derivation and thus facilitates language acquisition. Similarly, other researchers have argued that sound-shape mapping biases depend upon physical experience with one’s own vocal tract gestures, tongue position or lip shape when producing speech sounds (Berlin, 2006; Maurer et al., 2006; Ohtake & Haryu, 2013). Patterson and Werker (2003) found that infants as young as 2 months of age can match a vowel with the particular face when the vowel was produced. They showed the infants two faces of the same person (either a female or a male) when articulating /a/ and /i/ simultaneously side by side, and one sound track of a vowel (either /a/ or /i/) was played repeatedly. Performance was significantly above chance, and comparable to that of 4.5-month-old infants’ percentage of total looking time

(63.7%) using the same stimuli (Patterson & Werker, 1999). Detecting this kind of adult-like ability (Green & Kuhl, 1989) in 2-month-old infants, Patterson and Werker argued that speech information is quite special, and remarkable since infants are not able to match other kinds of information such as emotion, age, gender, etc. until at least 6 months of age (Walker, 1982; Bahrick, Netto, & Hernandez-Reif, 1998; Patterson & Werker, 2002). They also argued that speech-related crossmodal correspondence such as mappings between vowel sounds and lip shapes need no, or very rapid learning.

Although this idea that lip rounding is the main driver of sound symbolic effects seems to make intuitive sense, it should be noted that most language users are not explicitly aware of the articulatory dynamics of their vocal tracts during speech, and at this stage there is little direct evidence to support this hypothesis. Moreover, preliminary data from Styles (2014) suggested that /y/ shares tongue height and tongue frontness with /i/ and tongue height and lip rounding with /u/, but sound-shape mappings only differed significantly from /u/, suggesting vowel backness is a stronger driver of sound symbolic effects on vowels than lip rounding.

Although the mechanisms underlying sound symbolism remain unclear, many theorists agree that it seems to play a useful role in early language development (Imai et al., 2008; Imai & Kita, 2014; Maurer et al., 2006; Monaghan, Mattock, & Walker, 2012; Peña, Mehler, & Nespor, 2011) and

possibly language evolution (Berlin, 2006; Imai & Kita, 2014; Ramachandran and Hubbard, 2001).

1.8 ERP and fMRI Evidence for sound symbolism

Other researchers are investigating the neural substrates of the bouba-kiki effect using neurophysiological methodologies. In a category judgement task following novel spoken word learning, Kovic, Plunkett, and Westermann (2010) reported that within 200 ms after the onset of a visual stimulus, Event Related Potentials (ERPs) were more negative in the sound-symbolic condition than in the non-sound-symbolic condition over the occipital region. They interpreted this stronger negativity in the sound-symbolic condition as an early auditory-visual integration process. More recently, in a lexical decision task for written stimuli presented inside shapes, Sučević, Savić, Popović, Styles, and Ković (2015) detected stronger negative deflections following phonologically incongruent stimuli than for congruent stimuli in a similar, early time window (100-160ms). The findings suggest that the effects appeared very fast after the onset of a stimulus.

Arata, Imai, Kita, Thierry, and Okada (2010) simultaneously presented 12-month-old infants with a bouba-kiki-type figure (either round or spiky) and an auditory pseudoword (either “kipi” or “moma”), where the shape and the sound were sound-symbolically matched (match condition) or mismatched (mismatch condition). The infants showed different wave patterns across the two

conditions after about 200 ms of the sound onset, which may be the P2. P2 responses were suggested to be linked to multisensory integration (Bien, ten Oever, Goebel, & Sack, 2012) and influenced by sound-symbolism (Lockwood & Tuomainen, 2015). Arata et al. (2010) argued that the results that infants are sensitive to audiovisual correspondence support the neonatal synaesthesia hypothesis (Maurer, 1993; Maurer & Mondloch, 2005) which claims that infants are like synesthetes, showing more cortical connectivity than average adults, resulting in their ability to detect sound symbolism. Using a similar paradigm, except that a curvy or spiky shape was presented prior to a verbally presented sound (“kipi” or “moma”), Asano et al. (2015) observed increases in amplitude in the gamma band in the sound-shape congruent condition within 300 milliseconds in 11-month-old infants. Moreover, they observed greater N400 responses, an indicator of semantic integration difficulty, in the incongruent condition compared to the congruent condition, suggesting the extra effort was required to integrate incongruent cross-modal information. They further argue that congruent sound symbolic information may help infants to map speech sounds to referents and thus facilitate early word learning (Imai et al., 2008; Maurer et al., 2006). Recently in Lockwood, Hagoort, and Dingemanse’s (2016) study, Dutch participants learned Japanese ideophones with translations of either real meanings or the opposite meanings. Participants were significantly better at identifying the words they learned with real meanings than those with opposite meanings. The analysis of their ERP data showed that the Japanese ideophones learned in the real condition elicited a

greater P3 component and late positive complex than those learned in the opposite condition. These findings lend new behavioural and electrophysiological support to the idea that sound symbolism facilitates word learning.

Moreover, a recent fMRI study (Revill, Namy, DeFife, & Nygaard, 2014) found increased activation for sound-symbolic spoken foreign words compared to non-symbolic ones in the left superior parietal cortex in a meaning-matching task. They also found a positive correlation between accuracy on sound symbolic word guessing and fractional anisotropy in the left superior longitudinal fasciculus. They pointed out that the observed activated areas are involved in cross-modal processing, providing evidence that sound symbolism in language is related to general cross-modal processing. These results also lend support to cross-linguistic/universal sound symbolism rather than language-specific sound symbolism (see Imai & Kita, 2014, p5 for more discussion on universal vs. language-specific sound symbolism).

In total, electrophysiological and neurophysiological measures provided more evidence for the existence of sound symbolism and that sensitivity to sound symbolism can facilitate word learning in both children and adults.

1.9 Developmental disorders & poor sound-symbolic performance

Occelli, Esposito, Venuti, Arduino, and Zampini (2013) found high functioning autistic participants performed significantly worse than normal controls at bouba-kiki paradigm, despite performing at above-chance level, and low functioning autistic participants were at chance level. They argued that such results are due to a deficiency in multisensory integration in autistic participants, and they further linked the cross-modal correspondence effect to integration process between motor and sensory areas in the left inferior frontal gyrus. However, they did not test other kinds of crossmodal correspondence tasks in those autistic individuals, nor did they examine their multisensory integration impairment. Thus, it is inappropriate to conclude that the failures of typical crossmodal correspondence in autistic subjects are due to an impairment in multisensory integration in the same individuals.

Similarly, people with dyslexia also have trouble with cross-modal mappings. Drijvers, Zaadnoordijk, and Dingemanse (2015) investigated sound symbolism with Dutch dyslexic participants. Based on results of previous work (e.g., Nielsen & Rendall. 2011; 2012; Maurer et al., 2006), Drijvers et al. created ‘rounded’ stimuli using sonorants (/m/, /n/, /l/, /r/) and rounded vowels (/u/ and /o/); ‘pointy’ stimuli using obstruents (/z/, /g/, /v/, /b/) and unrounded vowels (/i/ and /e/), ending up with twenty CVCV-type pseudowords, e.g., pointy: /zige/, /gebi/, /vebi/, /bive/; rounded: /mumo/, /lonu/, /nuro/, /romu/. The visual stimuli consisted of 100 randomly generated single Gestalt shape figures

(50 round ones, 50 jagged ones). In each trial, one pointy and one rounded randomly selected figures were presented side by side for two seconds first, and then participants were asked which of the two figures was a certain auditory stimulus (e.g., /bege/). As what the researchers expected, people with dyslexia made significantly fewer valid sound-shape mappings than controls.

Considering the complex nature of dyslexia, it is unfair to say the deficit in cross-modal abstraction process alone causes dyslexics to be worse at sound-symbolic associations. Previous studies (Ramachandran, 2004; Ramachandran & Hubbard, 2001) suggest that the angular gyrus is involved in sound-symbolism processing, and the dyslexic population have shown lower activation and connectivity patterns in this area. More neurobiological evidence is needed to reveal the relationship between dyslexia and audiovisual multisensory perception. Moreover, the auditory stimuli were less systematically selected in this study, in particular, according to D'Onofrio (2013), /b/ and /g/ are consistently matched with round shapes regardless of whether followed by rounded vowels (/u/, /a/) or unrounded (/i/, /e/) vowels, and thus considered as drivers of the sound-symbolism effects. A bigger difference in sound-symbolic effects between dyslexics and normal controls may be expected when using more systematically contrastive stimuli.

So far, this section has reviewed evidence for the phenomenon of sound symbolism in the language domain using different measures (i.e., behavioural, electrophysiological and neurophysiological) cross-culturally in normal participants and people with developmental disorders from different age groups.

In the next section, we will review some research on non-lexical-pitch related cross-modal correspondences outside of language.

1.10 Systematicity outside language: pitch effects

Outside of the language domain, cross-modal congruence between auditory pitch and other sensory modalities has been widely investigated (see Spence & Deroy, 2013 for a review). For example, when people draw visual imagery for short musical selections they hear, they tend to draw thinner, brighter, smaller and more angular lines or forms for music higher in pitch (Karwoski, Odbert, & Osgood, 1942). Some early empirical evidence of pitch-related cross-modal correspondences comes from Marks' laboratory experiments. For example, when given a set of colors varying in lightness and a set of notes varying in pitch, participants consistently paired the higher pitch with the lighter color. So the relationship between color brightness or lightness and auditory pitch is the higher the pitch of the sound, the lighter or brighter the color of the image (Marks, 1978). He also found cross-modal correspondences between pitch and direction (up/down) as well as pitch and sharpness (Marks, 1974; 1987). Marks (1987) asked participants to match two auditory stimuli (one 220-Hz saw-tooth wave and one 360-Hz saw-tooth wave) to two visual forms (one U-shape and one V-shaped respectively) and found that the high pitch sound was matched to the angular shape and the low pitch sound was matched to the round shape. Moreover, when asked to adjust pure tone frequencies to best fit visual stimuli,

participants more often assigned lower frequencies to round figures than to other figures (O'Boyle & Tarte, 1980). Marks (1978) interpreted the observed audiovisual correspondences as 'weak synesthesia' which resembles strong synesthesia found in synesthetes. He further argued that both types of synesthesia are about how stimuli from different sensory modalities are related and both types are systematic, but that cross-modal correspondences in weak synesthesia are contextual, while cross-modal correspondences in strong synesthesia are absolute (one-to-one mapping).

Walker and Smith (1985) asked subjects to press different keys differing in size depending on which of four different words they heard and each word was accompanied by a high pitch tone (5500 Hz) or a low pitch tone (50 Hz). The subjects' responses were slower when the pitch of the sound and the size of the key on which they responded were incongruent (e.g., high pitch tone –big size key or low pitch tone—small size key). They argued that the results suggest that multisensory sensitivity takes place even when explicit sensory cue is absent and thus automatic.

Moreover, in another attempt to investigate the relationship between auditory pitch and visual size, Gallace and Spence (2006) asked participants to judge relative visual sizes of objects accompanied by task-irrelevant sounds, and found faster responses to congruent trials (e.g., a big disk with a low pitch sound) than to incongruent ones (e.g., a bigger disk with a high pitch sound),

lending support to the crossmodal correspondence between auditory pitch and object size.

Ward, Huckstep, and Tsakanikos (2006) found sound-colour synaesthetes tended to see lighter color for higher pitch sounds and non-synaesthete controls did the same thing by intuition. Thus, they speculate that to some extent synaesthesia and normal cross-modal perception share the same cognitive mechanism. However, the mechanism differs in terms of precision and automaticity. They further speculate that synaesthetes' mappings between sounds and colours are precise and automatic, while non-synaesthetes' mappings are noisy and strategic rather than automatic.

Speeded classification tasks have often been used in research on cross-modal correspondences, because people tend to respond to a stimulus accompanied by a task-irrelevant stimulus in other sensory modalities that is cross-modally congruent with it faster and more accurately than incongruent ones. Using speeded classification paradigm, Evans and Treisman (2010) found cross-modal relationship between auditory pitch and visual position, size, and spatial frequency. Namely, high pitch is cross-modally related to high visual position, small size and high spatial frequency. They also argue that the processing of the irrelevant dimension is automatic and the founded audiovisual correspondence may be at a low perceptual level.

To rule out the uncertainty of how much of the cross-modal correspondence effect is due to the presence of an irrelevant stimulus and how much is due the

stimulus compatibility, Parise and Spence (2012) used a modified version of the Implicit Association Test (IAT), where a certain visual stimulus and a certain auditory stimulus are mapped to one response key while the other visual stimulus and auditory stimulus are mapped to the other response key and only one stimulus is presented at a time. Parise and Spence (2012)'s results replicate Köhler's *takete/maluma* paradigm (*takete*-pointy; *maluma*-curvy), Sapir's *mil/mal* paradigm (*mil*-small; *mal*-big) audiovisual correspondence between auditory pitch and size (high pitch sound and small size; low pitch sound and big size) and documented two new audiovisual cross-modal correspondences (high/low pitch sound-acute/obtuse angle; square/sine waveform tones-angles/curves). Due to the presence of a RT modulation faster than 400ms, they argued that the cross-modal congruence effect might be automatic.

Toddlers' and preverbal infants' pitch-related cross-modal correspondences have also been assessed. Mondloch and Maurer (2004) investigated pitch-brightness correspondence and pitch-size correspondence in young children (30-36 months). In one experiment, they asked the children to decide which of the two bouncing balls (one small and white, one large and gray) made which of the two sounds (one high-pitched, one low-pitched). To explore pitch-brightness correspondence and pitch-size correspondence separately, in a second experiment, they used balls varying only in surface brightness, and the balls used in a third experiment differed in size only. The great majority (11 out of 12 children) matched the small, white ball with the high-pitched sound and the large, grey ball with the low-pitched sound. Similarly, all the children

matched the white ball with the high-pitched sound and the grey ball with the low-pitched sound. Similarly, most of the children (9 out of 12 in the first group, 10 out of 12 in the second group) matched the large ball with the low-pitched sound and the small ball with the high-pitched sound. Mondloch and Maurer (2004) claimed that such observations might be due to the remnants of neonatal cross-modal connections in the brain.

Haryu and Kajikawa (2012) examine 10-month-old infants' sensitivity to cross-modal correspondence using a violation-of -expectation procedure. In the pitch-brightness experiment, the infants showed significantly longer looking time for the incongruent condition where a light gray object bounced with a low-pitched tone than for the congruent condition where a light object bounced with a high-pitched tone. The authors proposed that such results were because the incongruent condition was more surprising to the infants. On the other hand, in the pitch-size experiment, the infants looked at the both conditions equally long, failing to detect their sensitivity to pitch-size correspondence. Hence, Haryu and Kajikawa (2012) argue that the infants' sensitivity to pitch-brightness correspondence is somewhat initial (innate), while their sensitivity to pitch-size correspondence may be learned rather than innate. Walker et al. (2010) also argue that after birth infants' encounters of co-occurrences of pitch and size may be sufficient to enable them to acquire the cross-modal correspondence between pitch and size.

Pitch-related cross-modal correspondences have been confirmed in even younger infants. Walker et al. (2010) examined four-month-old infants' sensitivity to cross-modal correspondence between pitch and visuospatial height and correspondence between pitch and visual sharpness. In a pitch-visuospatial height experiment, a ball was moving vertically accompanied by a sound which changed in pitch. In the congruent condition, the ball rose accompanied by an ascending tone and fell with a descending tone, and vice versa for the incongruent condition. In the second experiment which investigated pitch-shape correspondence, a shape became more and less pointed accompanied by a sound varying in pitch. In the congruent condition, the pitch of the sound rose and fell as the shape became more and less pointed, and vice versa for the incongruent condition. In both experiments infants looked longer at congruent trials than incongruent ones. These results demonstrate that sensitivity to correspondence between auditory pitch and visual spatial height as well as correspondence between auditory pitch and visual sharpness was found within first 3 or 4 months of life. Walker et al. (2010) further argue that such findings provide evidence for the idea that neonatal perception is synaesthetic (Maurer, 1993; Maurer & Mondloch, 2005).

Dolscheid, Hunnius, Casasanto, and Majid (2014) noted that pitch is described differently in different languages. For example pitch can be described as "high" or "low" in English, while in some languages, pitch is "thin" or "thick". They examined 4-month-old Dutch infants' sensitivity to pitch-height mappings and pitch-thickness mappings using a preferential-looking paradigm.

The infants looked significantly longer at cross-modally congruent trials for both types of mappings, which indicated that different from language-specific adults, infants are sensitive to the both variants of space-pitch correspondence. Dolscheid et al. (2014) also argue that space-pitch correspondences originate from preexisting mappings rather than language. Language, however, changes these associations gradually via competitive associative learning.

Fernández-Prieto, Navarra, and Rons (2015) also investigated infants' sensitivity to cross-modal correspondences between pitch and size using dynamic auditory and visual stimuli. Cross-modal correspondence effects were found in 6-month-old infants but not in younger infants. Agreeing with Dolscheid et al. (2014), Fernández-Prieto et al. claimed that language is not necessary for multisensory perception to take place. They further speculated that sensitivity to pitch-size correspondence may initially occur during the first month after birth and is acquired at around 6 months of age. The development of this sensitivity is influenced by the environment including language experience. They also point out that Haryu and Kajikawa (2012)'s failure to detect 10-month-old infants' sensitivity to pitch-size correspondence may be due to the asymmetry between their auditory and visual stimuli. That is, their auditory stimuli were static whereas the visual stimuli were moving. This asymmetry between stimuli may lead to their failure to establish links between auditory and visual stimuli.

Moreover, Bien et al. (2012) investigated adults' sensitivity to pitch-size correspondence using a combination of psychophysics and ERPs. They found audiovisual congruent conditions (e.g., high pitch-small size) elicited a greater P2 response at around 250 ms in intraparietal areas which is consistent with Asano et al. (2015)'s findings in infants. Bien et al. (2012) argue that the P2 response is involved in multisensory integration. Hence, in general, neural findings from both adult and infant support the idea that P2 responses are related to multisensory perception.

Sensitivity to cross-modal correspondence is not limited to humans. When chimpanzees were required to classify squares in terms of color (white; black) while hearing task-irrelevant high-pitched or low-pitched sounds, they performed better in cross-modally congruent conditions (i.e., high pitch sounds and white squares; low pitch sounds and black squares) than incongruent conditions (high pitch sounds and black squares; low pitch sounds and white squares), suggesting such natural mappings might also be present in our nonlinguistic ancestors and may have influenced language evolution (Ludwig, Adachi, & Matsuzawa, 2011) .

To date, only non-lexical auditory pitch has been investigated in the context of crossmodal correspondence. To the best of our knowledge, lexical tones have not yet been systematically studied in the context of crossmodal correspondence. This is the first work that systematically investigates the

crossmodal correspondences between lexical tones (i.e., Mandarin Chinese tones) and some visual dimensions (i.e., visual angularity, spatial frequency).

Some theories address physical mechanisms underpinning the links between auditory pitch and vision. According to fundamental frequency theory, vocal sounds signal physical properties such as size, shape and dominance of the speaker. For mammals, the larger the sound production apparatus (i.e., vocal-tract), the lower the fundamental frequency (F_0). For example, adults and males have larger vocal tracts than children and females, resulting in lower pitched voice for the former, than for the latter (Morton, 1994; Ohala, 1994; Thompson & Estes, 2011). The Frequency Code hypothesis suggests that “high tones, vowels with high second formants (e.g., /i/), and high-frequency consonants are associated with high-frequency sounds, small size, sharpness, and rapid movement; low tones, vowels with low second formants (e.g., /u/), and low-frequency consonants are associated with low-frequency sounds, large size, softness, and heavy, slow movement” (Ohala, 2006). The second formant (F_2) is correlated with vowel backness and front vowels (e.g., /i/, /e/) are high in F_2 . Many studies have confirmed that the front vowel /i/ is associated with smallness (e.g., Sapir, 1929; Tarte & Barritt, 1971; Tarte 1974; Peña et al. 2011), while /u/ is chosen more often for curvy shape than large size (e.g., Köhler 1929:1947; Ramachandran and Hubbard, 2001; Maurer et al., 2006; Ozturk et al., 2013). Rather, another back vowel /a/ is more often picked for large size. For consonants, some studies (Ultan, 1978; Ohala, 1994) demonstrate that palatal consonants are high in pitch, and thus are related to

smallness. This argument has also been confirmed by Haynie et al. (2014), where they investigated 120 Australian languages and found that palatal consonants occur more frequently in words meaning “small” or “near” while velar consonants occur more often in words meaning “large” or “far”, which is consistent with the Frequency Code hypothesis. Evidence for the link between palatal consonants and smallness meaning can also be found in the language of Native American Karuk (Ultan, 1978).

Furthermore, Walker and colleagues (2010) attribute some pitch-shape correspondences to the hardness of natural materials. For example, sharp and pointed objects tend to be formed from hard materials, which in turn tend to produce high-pitched sounds when struck. Spence (2011) therefore argues that many cross-modal correspondences may be an outcome of learning about the multisensory environment. These two perspectives highlight different possible sources for multisensory processing. One model highlights our evolutionary heritage, which may be an innate source of universal linguistic biases, and the other highlights environmental exposure to the physical properties of objects.

In summary, cross-modal correspondences have been documented between non-linguistic pitch (pure tone) and visual size, angularity, spatial height, spatial frequency, etc. Linguistic uses of pitch, however, have not been systematically investigated in the cross-modal correspondence literature.

1.11 Linguistic experience tunes speech perception

Kuhl (2004) argues that infants are initially able to distinguish almost all phonetic units. In other words, Infants, ‘citizens of the world’, in their first few months can hear any sound in any language that contrasts. For example, the liquid contrast between /r/ and /l/ is perceived differently by American adults and Japanese adults. Namely, Japanese adults are unable to discriminate the /r/ and /l/ sounds in English. However, using a headturn task, Kuhl et al. (1997) found that at 6-8 months of age, the Japanese infants performed as well as the American infants by responding to a novel stimulus with a head-turn. Surprisingly, two months later, the American infants became much better while the Japanese infants became much worse doing the same task. Kuhl argues that newborns are universally able to discriminate between differences in all phonetic units. As they listen to ambient speech production, they take statistics on the language that they hear. The statistics of Japanese and English are very different. English has a lot of /r/s and /l/s. The distribution of Japanese is totally different, with a group of subtly different sounds which is known as Japanese /r/, which is characterized by a retroflex tap, rather than an approximant. Babies absorb the statistics of the language(s) they are exposed to and it changes their neural representations of sounds, changing them from the ‘citizens of the world’ to ‘culture-bound’ listeners. As adults, we are probably less efficient at taking on those statistics due to the neural commitment that occurs during childhood

(Kuhl 2004). We are heavily biased by the representations in memory that were formed early in development.

The change in phonetic perception from universal to language-specific happens to vowel sounds as well. Using the head-turn technique, Kuhl et al. (1992) tested 6-month-old infants in the United States and Sweden on two vowels. One is the American English /i/, a front unrounded vowel in the word *feet*; the other is the Swedish /y/, a front rounded vowel in the Swedish word *fy*. Using computer they synthesized prototypes of the American English /i/ and Swedish /y/. They then created 32 variants similar to each prototype for each prototype. Each group (English or Swedish) of infants heard a repeating prototype vowel (American English /i/ or Swedish /y/). The infants were trained to produce a head-turn when the prototype vowel was changed to one of its variants. The infants detected more variants of the prototype vowel for the native language compared with the foreign language, indicating language experience during the first few months after birth reinforces native-language phonetic perception while undermining the non-native phonetic perception.

Is the perceptual reorganization for phonemes also true for lexical tones? Fundamental frequency (pitch) variation is perceived differently between tone language learners and non-tone learners (Gandour, 1983; Jongman, et al., 2006; Singh, Hui, Chan, and Golinkoff, 2014). When do tone-language-exposed infants start recognizing pitch changes distinguish word meaning in their native language and showing initial perceptual sensitivity to lexical tones? When do

non-tone language infants begin ignoring pitch changes that do not cause difference in word meaning in their language?

Mattock and Burnham (2006) used the head-turn procedure to investigate tone perception in 6- and 9-month-old tone-learning Chinese infants and non-tone-learning English infants with Thai tone contrasts, along with non-speech tone contrasts matched for F0 movement without phonetic cues. Chinese infants' performance on both lexical and non-lexical tones are equally good at 6 and 9 months. By contrast, English infants did significantly worse in the same task at 9 months compared to 6 months, whereas their non-linguistic tone discrimination remained unchanged over age. These results demonstrate that perceptual reorganization for tones is language specific. Also, given that the Chinese infants processed the nonnative tones (i.e., Thai tones), their maintenance of tone perception over age serves as strong evidence for effect of linguistic experience on early tone perception development.

Mattock, Molnar, Polka, and Burnham (2008) tested 4-, 6-, and 9-month-old infants from monolingual English and monolingual French families using Thai tones. There was tone discrimination decline between 6 and 9 months and not before. Yeung, Chen, and Werker (2013) replicated Mattock et al. (2008) by finding that English-hearing infants showed similar declines in discrimination of Cantonese tones from 4 to 9 months of age. Moreover, Liu and Kager (2014) also replicated deterioration in tonal (i.e., Mandarin Chinese tones) sensitivity

in Dutch infants at around 9 months, whereas their tone perception rebounded at 17-18 months, suggesting perceptual flexibility in infancy.

Although the innate capacity of distinguishing phonetic units is universal, we start accommodating to our native language system from a few months after birth. Our linguistic experience rapidly makes us understand what kinds of phonetic contrasts are statistically well-represented in our language. We are subsequently biased towards native-language-specific phonetic units. For tone perception, the shift begins at around 6 months of age, after which tone-learning children keep developing their perception and production of tones while non-tone-learning children's tone perception deteriorates as tones are less represented in their languages. Accordingly, we expect different responses to tone-related multisensory perception tasks between tone-language-speaking and non-tone-language-speaking participants.

1.12 Pitch in language

Vowels and consonants can be described by acoustic properties, such as formant frequency values and transitions. Lexical tones refer to syllable-level variations in the temporal contour of the fundamental frequency (F0) (or pitch) and serve to draw contrasts between word meaning (Singh & Fu, 2016; Jongman et al., 2006; Gandour, 1978). Mandarin Chinese is a tone language and has four lexical tones. In terms of pitch, Tone 1 has high-level pitch, Tone 2 mid-rising, Tone 3 low-dipping (or falling-rising) and Tone 4 high-falling

(Chao, 1948) (see Figure 1.1). As mentioned, just as vowels and consonants, tones are involved in distinguishing word meaning. For example, when the syllable /yan/ is produced in the four tones, each word has a distinct meaning: 烟 (yan₁, smoke); 盐 (yan₂, salt); 眼 (yan₃, eye); 艳 (yan₄, colourful). As seen, therefore, tone information is vital in word discrimination in Mandarin Chinese. In addition to pitch movement, other factors including duration, amplitude, tonal coarticulation, and speaker's F0 range can also influence the perception of Mandarin tones (Jongman et al., 2006; Singh & Fu, 2016). Tone 2 and Tone 3 tend to be the longest, and Tone 4 the shortest (Jongman et al., 2006). Tone 4 is also the highest in overall amplitude (Chuang & Hiki, 1972), changing most in pitch, while Tone 1 is very stable and does not change much in pitch. In one study, Tone 3 was perceived as a low, unchanging pitch and subjects had difficulty distinguish between Tone 2 and Tone 3 (Whalen & Xu, 1992).

Figure 1.1. Time by frequency representations of [u] uttered in Mandarin Chinese tones by a female native Mandarin speaker (Tone 1: black line; Tone 2: yellow line; Tone 3: blue line; Tone 4: red line) (see p.44 & p52 in the following Chapter for details).

In non-tone languages, fundamental frequency (pitch) variation is non-phonemic while formant height and contour are phonemic (Braun & Johnson, 2011; William, 1977). On the other hand, in tone languages, fundamental frequency variation can be either phonemic (tone) or non-phonemic (i.e. intonation) (Gandour, 1978) which are functionally different, despite some structural overlap (Singh et al., 2014). Therefore, the acquisition of lexical tone is relatively complicated. However, lexical tone acquisition has been thought to be completed at a rapid rate and within a quite short period of time (Li & Thompson, 1977). Previous studies reveal that Mandarin and Cantonese infants as young as four-to-six months showed sensitivity to native or foreign lexical tone contrasts (Thai tone contrasts; Cantonese tone contrasts) (Mattock & Burnham, 2006; Yeung et al., 2013).

When considering the acquisition of the different tones, a number of previous studies have found that Tone 1 and Tone 4 tend to be acquired prior to Tone 2 and Tone 3. Children are often confused by Tone 2 and Tone 3 and substitute them for each other throughout tone acquisition process. Moreover, by the age of three, Mandarin learning children are able to perceive three of the four lexical tones in monosyllabic words, except for Tone 3. In terms of both tone production and perception, Tone 3 seems to be acquired last. Tone 3, therefore, seems to be the most difficult one for both perception and production acquisition (Li & Thompson, 1977; Wong, Schwartz, & Jenkins, 2005; Lin, Wang, & Shu, 2013). Some studies (Lee & Nusbaum, 1993; Ho & Bryant, 1997; Siok & Fletcher, 2001; Tong, Francis, & Gandour, 2008; Lin et al., 2013) have

also found that vowels seem to have a more important role in tone processing than consonants.

Tong, Francis, and Gandour (2008), using a speeded classification paradigm, asymmetric interference effects were observed between segmental and suprasegmental dimensions, with segmental dimensions interfering more with tone classification than the reverse. Among the three dimensions, vowels exerted greater interference on consonants and tones than vice versa (Also see, Lee & Nusbaum, 1993). Moreover, both Cantonese and Mandarin speaking children are able to process tones alone in the absence of segmental variability in such as onsets and rimes after 5 years old, suggesting that the enhancement of metalinguistic awareness and literacy exposure may contribute to the tone processing development.

How about the perception of lexical tones for non-tone-language learners and speakers? A reduction in sensitivity to lexical tones (Thai tones; Cantonese tone) was found in non-tone-hearing infants (English infants and French infants) between 4 to 9 months of age (Mattock & Burnham, 2006; Mattock et al., 2008; Yeung et al., 2013) and this reduced sensitivity to lexical tone extends to adulthood (Tsushima et al., 1994). However, non-tone-language adults are still sensitive to lexical tones that they hear (Francis, Ciocca, Ma, & Fenn, 2008; Gandour et al., 2000; Halle, Chang, Best, 2004; Kaan, Barkley, Bao & Wayland, 2008; Xu, Gandour, & Francis, 2006). For English speakers, their Mandarin tone perception seems to be influenced by both their native

intonation system and the presentation position of Mandarin tones. Broselow, Hurtig, and Ringen, (1987) found that Mandarin Tone 4 was well detected when presented in monosyllables or at the end of two syllables or three syllables. On the other hand, Tone 4 became poorly identified in non-final positions. Both Mandarin Tone 4 and English unmarked ends of declaratives have a falling pitch. The authors therefore claim that Tone 4 is remarkably accurately perceived by the American listeners when it is presented in isolation or at the end of syllables due to the influence from their native intonation system.

Some research further demonstrates that non-tone language speakers' both perception and production of Mandarin tones can be largely improved by training (Leather, 1990; Wang, Spence, Jongman, & Sereno, 1999; Wang, Jongman, & Sereno, 2003). This improved tone sensitivity is generalizable and retrievable a few months after training and leads to a native-like cortical change in the brain (Wang et al., 2003), suggesting stable experience-dependent perceptual and production plasticity.

1.13 Sound symbolism in the Chinese language

To date, very little work has investigated the nature of sound symbolism in the Chinese language (Mandarin), and of those studies that exist, the majority have focused on properties of the language which are common in European languages (i.e., vowels), rather than the less familiar features (i.e., tones). For

example, Huang, Pratoomraj, and Johnson (1969) agreed with Johnson (1967) and Newman(1933) that, in English, sound-size correspondence is conveyed by initial vowels of words, and that words beginning with /o/ show the strongest association with largeness, getting weaker in the order /o/, /u/, /a/, /e/, /i/. Huang and the colleagues (1969) asked students from Taiwan to produce words meaning or associated with largeness or smallness. Participants produced more words whose initial vowels were /i/ associated with smallness (Smallness-Largeness ratio: 9:1) than largeness. On the other hand, they produced more words whose initial vowels were /a/ (Smallness-Largeness ratio: 0.59:1), /u/ (Smallness-Largeness ratio: 0.17:1) or /o/ (Smallness-Largeness ratio: 0.96:1) associated with largeness than smallness. They therefore argued that their Chinese language findings were consistent with those of English and Thai in the way that /i/ is associated with smallness, and that /a/, /u/ and /o/ are associated with largeness. However, since the words consisted of vowels and consonants, it is impossible to tease apart sound cues provided by both vowels and consonants. Furthermore, Chinese is a language with four lexical tones. Therefore, not only vowels and consonants, but also tones may influence sound symbolism in the Chinese language. The present body of work will explore cross-modal correspondence between Mandarin tones and different visual dimensions.

To the best of our knowledge, just one study has attempted to investigate sound symbolism for Chinese tones. In one of Lapolla's (1994) studies, he investigated whether native English speakers who had no knowledge of the

Chinese language were sensitive to sound-meaning correspondences embedded in Mandarin Chinese sound cues. He asked native English-speaking participants to choose one from an auditorily presented pair of Chinese words for a written English meaning. The results were not significant ($p < 0.30$), meaning the native English speakers were not sensitive to Chinese sound-meaning correspondence cued by Chinese words. He further attempted to investigate tone effects on this unproven sound-meaning correspondence by reversing tones on the words in those pairs whose members differed in tone. There were significantly more correct responses than incorrect ones this time ($p < 0.05$), when the tones were wrong, indicating cross-linguistic sound-meaning correspondence was *reversed* for speakers who were unfamiliar with the language, suggesting that people may have different sound-meaning mappings depending on their language experience. Similar to Huang et al. (1969), since Lapolla (1994) used nonsense whole words as auditory stimuli it is hard to determine the precise effect of vowels, consonants or tones on sound-meaning correspondences. In the current research, we will examine the relationship between Mandarin tones and two visual dimensions (i.e., visual angularity and spatial frequency) as well as the relationship between spoken vowels and these visual dimensions using single vowels using a more systematic approach. The results will also be discussed in terms of the possibility of different sound-meaning mappings due to different language backgrounds.

1.14 The current study—Mandarin-tone-related Sound Symbolism

Lexical tones exist in about 70% of the world's languages (Yip, 2002) and over half the world's population speak a tone language (Fromkin, 1978), but lexical-tone-related cross-modal correspondence has rarely been systematically studied, in particular, Mandarin Chinese tones. In the present study, we systematically explore sensitivity to cross-modal correspondences between visual angularity, spatial frequency, single vowels and Mandarin tones in different language groups (i.e. dominant Mandarin Chinese; balanced Mandarin-English; non-Chinese). Tones are documented to be more easily identified when presented in isolation (i.e. monosyllables) (Broselow et al., 1987). Confound effects might be caused by combinations of vowels and consonants. Therefore we use single vowels ([i] & [u]) articulated in Mandarin tones.

In previous research, for the tone contrasts, researchers with different language backgrounds tended to have different selections in the literature. When Chandrasekaran, Krishnan, and Gandour (2007) try to compare mismatch negativity (MMN) response to pitch movements of Mandarin Chinese tones in Chinese and English subjects, they compare two conditions in terms of degree of acoustic similarity, with Tone 2 and Tone 3 as the similar pair and Tone 1 and Tone 3 as the dissimilar pair. They find larger mean MMN amplitude in the Chinese group than the English group for the Tone1-Tone3 condition, suggesting language experience and perceptual saliency of tone influence the

early stage of lexical tone perception. Singh et al., (under review) also use Tone 2 and Tone 3 as the acoustically similar pair and Tone 1 and Tone 3 as the acoustically dissimilar pair to investigate Mandarin-English bilingual infants' sensitivity to Mandarin tones when listening to Chinese and English speech and find stronger effect for Tone3 and Tone 1 condition as compared to Tone 3 and Tone 2 condition in both language speech. On the other hand, Liu and Kager (2014) examined Dutch infants' capacity to discriminate the Mandarin Tone1-Tone 4 contrast and find a U-shaped trajectory in non-tone-learning infants' tone perception development. That is, the infants have tonal sensitivity at 5-6 months, fail to show it at around 9 months and show it again at 17-18 months. They did not clarify the reason for their selection of T1-T4 contrast. However, as seen in Table 1.1 and Figure 1.1, in terms of average pitch (F0) height, Tone 1 is the highest and Tone 3 is the lowest, and thus the T1-T3 is the most salient contrast. However, if one considers pitch change, Tone 1 changes the least and Tone 4 changes the most, and thus the T1-T4 may be the most salient contrast. Gandour (1983) finds that English speakers pay attention almost only to F0 height of stimuli. Moreover, as pitch changes do not distinguish word meaning in English, the non-native Chinese speaking researchers, the Gandour group (Chandrasekaran et al., 2007) and Singh et al., (under review) may have been influenced by their own perceptual systems, in their choice of choose T1-T3 as the acoustically dissimilar contrast. On the other hand, for tone language speakers, pitch variations make difference in word meaning, therefore, Chinese native speaking researchers, for example, Liu (Liu & Kager, 2014) selects the

T1-T4 contrast rather than T1-T3 contrast. Although these findings can act as a guide for what we expect to see, we will base our stimulus selections on the results of Study 1. Our procedure starts with a systematic evaluation of sound symbolism for different tones in different groups.

Table 1.1. Average F0 values (in Hz) of [u] and [i] produced in the four Mandarin tones by a female native Mandarin speaker

Average F0 values		Average F0 values	
u1	286.35	i1	294.58
u2	229.48	i2	229.25
u3	181.88	i3	167.02
u4	230.55	i4	237.63

Taken together, the present research aims to replicate the well-documented ‘*u*-curvy, *i*-pointy’ vowel-shape cross-modal correspondence (e.g., Köhler 1929:1947; Ramachandran and Hubbard, 2001; Maurer et al., 2006; D’Onofrio, 2013). Moreover, to our best knowledge, this program of research is the first attempt to systematically investigate the influence of Mandarin Chinese tones on different language groups’ sensitivity to audiovisual cross-modal correspondences. We expect different patterns of sensitivity to Mandarin-tone-related multisensory perception processing in different language groups. That is, Chinese-speaking participants (C & C/E) would respond to Tone1 (T1) and

Tone 4 (T4) differently because T1 is the steadiest tone and T4 is the most changing one in terms of pitch change and Chinese speakers are sensitive to pitch change (Jongman, et al., 2006; Singh, et al., 2014). On the other hand, English speakers (E) will show different responses to T1 (high-pitched) and T3 (low-pitched), as English speakers are sensitive to pitch height (Gandour, 1983).

1.15. Analytical Methodology

Analysis of Reaction Time

This section reviews some common analytical methods for reaction time data. Reaction time (RT) is a common measurement in psychology research. Many researchers analyze RT data by conducting an analyses using parametric statistics on the sample mean (Van Zandt, 2002). However, RT distributions are typically not Gaussian (normal) but rather rise rapidly on the left and have a long tail on the right (for example, see Figure 1.2). This means that most of the responses are quite fast, and the slower the response, the fewer the data points observed. Hence, raw RT data are typically skewed, and contain outliers.

Figure 1.2. An example of distribution of reaction time in a speeded decisions task (data from Study 2A, English group in the present thesis)

Many common statistical procedures such as *t*-tests and ANOVAs assume that the data have a normal distribution. To normalize RT distributions, researchers typically transform the data and/or delete some extreme trials. Transformations can include the inverse or the logarithmic transformation, depending on the underlying structure of the data. Many researchers agree that the inverse transformation normalizes the distribution, reduces the effect of slow outliers, and therefore generally maintains good power (e.g., Imam, 2006; Spencer & Chase, 1996; Greenwald, Nosek, & Banaji, 2003). In a linear scale, the difference between units is constant. The scale is reverse by inverse transformation, and the transformation also compresses the long tail and stretches the fast responses.

Another option is to perform an analysis of the whole RT distribution using the ex-Gaussian fit analysis. The ex-Gaussian distribution of RT is a convolution combining a Gaussian (normal) distribution and an exponential distribution (Luce, 1986). The ex-Gaussian distribution is specified through

three parameters: μ , σ and τ . μ and σ are the mean and the standard deviation of the normal distribution respectively. τ describes the mean and the standard deviation of the exponential distribution (Hwang Gu, Gau, & Tzang, 2013). Luce (1986) describes the ex-Gaussian function as a model for decision-making inside the temporal space, and therefore, a model which might describe different cognitive processes. This option is usually used when dealing with data with high variability and/or with a large amount of abnormally slow RTs (e.g., data obtained from old people or individuals with ADHD) (Navarro-Pardo, Navarro-Prados, Gamermann, & Moret-Tatay, 2013; Hwang Gu, Gau, & Tzang, 2013).

The median RT is reported as a central tendency parameter for non-symmetrical RT distributions when there is large variability among participants (Whelan, 2008). Ratcliff (1993) reported that using the median generally resulted in lower power than using inverse transformation or cutoffs.

Researchers should make a decision about which treatment of data is suitable for their data. For example, they should check the distribution and test the transformations. In the following sections, the distributions of data in the datasets of the present thesis will be discussed in terms of the methods that have been described.

Tests of Transformation

In the data recorded in the experiments reported in this thesis, the distributions of our data are roughly symmetrical after the inverse

transformation (for example, see Figure 1.3). As long as the distribution is roughly symmetrical, the mean will be an adequate measure of central tendency (c.f., Keppel, Saufley Jr., & Tokunaga, 1992; Sirkin, 2006; Miller, Daly, Wood, Roper, & Brooks, 1997).

Figure 1.3. An example of distribution of inverse RTs (data from Study 2A, English group)

Logarithmic transformation of RTs also normalized the distribution to some extent. However, the distributions of Log RTs are less symmetrical than the distribution of inverse RTs (for example, see Figure 1.3 & Figure 1.4). Since the inverse transformation generated more symmetrical distributions for the datasets in the present thesis, the inverse transformation was used throughout.

Figure 1.4. An example of distribution of Log RTs (data from Study 2A, English group)

Even though the inverse RT produces symmetrical distributions, a few outliers remain. Cutoffs to remove extreme outliers are often based on the standard deviation (Whelan, 2008). To trim extreme data in the current datasets, responses greater than three standard deviations above or below the mean for each participant were excluded. After the application of the cutoffs, distributions were checked again to ensure symmetry (for example, see Figure 1.5). The distributions were roughly symmetrical, and thus the mean of the transformed RTs would be an appropriate measure of central tendency, allowing parametric statistics to be used.

Figure 1.5. An example of distribution of inverse RTs after the application of the cutoff (data from Study 2A, English group)

While reporting the data on the transformed scale is sufficient, from Figure 1.3 and 1.5 it is clear that the inverse transformation does not produce readily understandable numeric output. Some researchers suggest it may be necessary or convenient to visualize effects and report means on the original millisecond scale using back-transformed data (Baayen & Milin, 2010).

Therefore, back-transformed RTs ($RT=1/\text{inverse RT}$) are used for visual presentation and reporting of means throughout. Since the inverse transformation and the cutoff based on the standard deviation are suitable for the current datasets, the RT data from the IAT studies (Ch3, Ch4, & Ch5) are analyzed using this treatment of data. The main findings were consistent with those in the original analyses.

The main range of the present datasets is 300-3000ms, which is consistent with the common RT range for the Implicit Association Test (IAT) (Greenwald, Nosek, & Banaji, 2003; Maison, Greenwald, & Bruin, 2001). Hence, the ex-Gaussian approach might not be suitable for the present datasets, as they lack long and thick right tails in the RT distributions (see Figure 1 for an example). In addition, in general, the tau parameter of ex-Gaussian distribution was used when group differences in means were not observed (e.g., Geurts et al., 2008; Leth-Steensen, Elbaz, & Douglas, 2000). However, group differences in means are detected in the present thesis (see Results and Discussion sections of the IAT experimental chapters (Ch 3, Ch 4 & Ch 5) for more details). Since the inverse transformation generated symmetrical data from which means can be used as a measure of central tendency, and differences are observed between means, neither the Ex-Gaussian nor the median approach was necessary.

RT analysis for the Implicit Association Test (IAT)

In the IAT, RTs are usually inverse-transformed or log-transformed and extreme data are excluded before computing means. The differences in means across conditions are then analyzed by using ANOVA or T-tests (Greenwald, McGhee, & Schwartz, 1998; Maison, Greenwald, & Bruin, 2001).

D measure

The *D* measure was developed for analysis of RT data in the contrasting conditions of the IAT. *D* divides the difference between two experimental treatment means by the standard deviation of all the RTs in the two treatments ($D = (M_1 - M_2) / SD$) (Nosek, Bar-Anan, Sriram, Axt & Greenwald, 2014). Using the standard deviation as a divisor adjusts differences between means for this effect of underlying variability. Greenwald and his colleagues (Greenwald, Nosek & Banaji, 2003; Nosek et al., 2014) argued that compared to the conventional measures (e.g., Mean, Median, Log, Reciprocal) for RT data in the IAT, *D* better reflects underlying association strengths, more powerfully evaluates relations between association strengths and other variables of interest, and better reveals individual differences that are due to association strengths rather than other variables.

In the current thesis, the *D* measure was used to test if the group differences in treatment effects observed in Study 2 using the inverse RT

measure would also be observed using the *D* measure. In each study (i.e., 2A, 2B, 2C, 2D), the means of the RTs in the two experimental treatments were calculated separately. The standard deviation of all the RTs in the two experimental treatments was computed. The *D* score was calculated accordingly for each participant in the three subject groups and then one-way between-subjects ANOVAs were run to examine the differences in mean *D* scores across subject groups. The results and discussion of the reanalysis of Study 2 using the *D* measure were updated at the end of Study 2 on page 136.

Chapter 2: Two-Alternative Forced Choice (2AFC) Tasks of Shapes and Vowels with Mandarin Chinese Tones

2.1 Introduction

As mentioned in Introduction, the ‘*u*-curvy, *i*-pointy’ vowel-shape cross-modal mappings (e.g., Köhler 1929:1947; Ramachandran and Hubbard, 2001; Maurer et al., 2006) and ‘high-pointy, low-curvy’ (non-lexical) pitch-shape relationship (e.g., Marks, 1974; 1987; Walker et al., 2010) have been widely documented in the literature. The main objective of the present chapter is to establish cross-modal correspondence between lexical tones (i.e., four Mandarin Chinese tones) and visual angularity using two-alternative forced choice (2AFC) tasks for speakers of different languages. We aimed to find out whether the ‘high-pointy, low-curvy’ response pattern also works for the relationship between lexical tones and visual angularity, and whether participants with different language background perform the tasks similarly or show different response patterns. To investigate the relationship between visual angularity and Mandarin tones, we conducted two online two-alternative forced choice (2AFC) tasks: Study 1A and Study 1B. In study 1A, participants were presented with two shapes differing in angularity and one spoken vowel (/i/ or /u/) articulated in one of the four Mandarin tones and asked to match the sound with one of the two shapes. This study aims to examine both the vowel effect and the lexical tone effect as well as language background influence on cross-modal correspondence between visual angularity and Mandarin tones. Study 1B

aims to further investigate this cross-modal correspondence by presenting participants with one shape (curvy or pointy) and a minimal pair of tonal contrast in which the vowel is controlled and asking the participants to match the shape with one of the sounds. The existing studies have investigated sound symbolism using similar sounds (i.e., nonsense words consist of vowels and consonants), whereas the present study is looking at single vowels articulated in Mandarin Chinese tones. To date, lexical tones have not been investigated in the literature on sound symbolism.

2.2 Study 1A: Two shapes with one sound

Methods

Participants

One hundred and fourteen volunteer participants (64 females), aged 18 to 57 years, took part in the present study, conducted using Qualtrics online survey software. Participants should be over 18 years of age and sign the Information & Consent sheet approved by the IRB of Nanyang Technological University online. We designed the study for three groups: a Chinese dominant bilingual group (C); a Chinese-English balanced bilingual group (C/E) and a group of English speakers with no Chinese (E). Due to the online nature of the task, no fixed limits were set for group size. The study was closed on a predetermined day shortly after 100 participants were recorded, and before

analysis was conducted. Fourteen participants did not fall into one of these groups, and were excluded. Therefore, there were one hundred valid participants in the present study (C: 45; C/E: 30; E: 25).

A brief survey asked participants if they are bilinguals of English and Chinese. All the participants in the C group and the C/E group reported themselves as bilinguals of English and Chinese. Participants were also asked to rate their proficiency in each of their languages and dialects on a five-point scale where one represents highest competence and five represents lowest (see Appendix 2.1). Table 2.1 shows different subject groups' self-reported Chinese and English proficiency levels.

Participants were also asked about where they live now and their residence history in different life stages. Most of the C Group were currently living in Mainland China or had been living in Mainland China until their graduation from colleges or universities. Most of the C/E Group were Singaporeans who grew up and are currently living in Singapore. Group E do not have any knowledge of the Chinese language and came from a variety of countries (e.g., Singapore, U.K., U.S.A., Australia, Germany, etc.). The experimental procedure was approved by the IRB of Nanyang Technological University.

Table 2.1. Self-reported language levels

Group	Language Level (Range)	
	Chinese	English
C	1	2-4
C/E	1-2	1-3
E	N/A	1-2

Stimuli

The visual stimuli were two ivory three-dimensional novel objects (one curvy; one pointy), photographed against a black background (see Figure 2.1). The auditory stimuli were the high front non-rounded vowel /i/ (the ‘ee’ vowel in ‘feet’) and the high back rounded vowel /u/ (the ‘oo’ vowel in ‘shoe’) articulated in the four tones of Mandarin Chinese by a female native speaker. The auditory stimuli were produced a minimum of three times each as isolated monosyllables, with the syllables arranged in a number of different orders, to maximize variation across the recorded set of stimuli. The auditory stimuli were recorded in a sound-proof recording lab using a Shure SN81 microphone and Acoustica recording software. The recording was done at a 44.1 kHz sampling rate with 16 bit encoding. Audio were edited and trimmed using GoldWave software.

Figure 2.1. Visual stimuli of Study 1A

Auditory stimulus validation

To ensure that the audio tokens were sufficiently standard for use in the test, we asked seven bilingual speakers of Mandarin Chinese and English to evaluate the typicality of each recording as an exemplar of the tone produced on the vowel in question. We played each sound, and asked what vowel it was, what tone it was, and asked people to rate the typicality of each sound, on a scale from one to seven. Seven represented the most typical and one represented the least typical. People could also use zero, if they thought it did not sound like the category at all. All seven raters agreed on the identity of the vowels and the tones. They rated the stimuli as very typical, with median scores of 5 or 6 for all items (see Table 2.2). Figure 2.2 shows the pitch tracks of the 8 auditory stimuli, where it is clear that the /i/ and /u/ track within each tone are more similar than the pitch tracks between tones. That is to say, each tone was clearly differentiated from the others, and was consistent across the two vowel tokens. Compared to auditory stimuli in previous studies (Moore & Jongman, 1997; So & Best, 2010; Whalen & Xu, 1992; Xu, 1997), T3 is not longer than the other tones. T1 has a fall-rise curve near the end. There is a slight rise at the end of T4, which is effectively inaudible, as it occurs in the quietest part of the articulation. The differences in the pitch contour between the present study and the previous studies might be due to the individual differences in speakers' vocalization features.

Table 2.2. Rating of typicality of the auditory stimuli

Sounds	Median	Range
i1	6	5-7
i2	6	3-7
i3	6	4-7
i4	5	4-7
u1	5	4-7
u2	5	3-7
u3	6	4-7
u4	5	3-7

Figure 2.2. Pitch tracks of the eight auditory stimuli used in Study 1A ([i] stimuli: red lines; [u] stimuli: black lines). The figure was produced using PRAAT.

Experimental platform

The experiment was presented using Qualtrics online survey software. Participants were instructed to run the experiment using laptops or computers in a quiet environmental background using headphones, since the platform did not guarantee stable audio on mobile devices.

Procedure

In the online procedure, after obtaining consent and adjusting audio volume, participants were presented with two large pictures side-by-side and a button which triggered the audio file to play. Each sound file was 500 ms long. In each question, participants could listen to the auditory stimulus as many times as they liked and they were asked to decide which of the two visual stimuli was a better match for the sound. Once a decision was made, they were instructed to move on to the next question by pressing the '>>' button, and each page was unreturnable. The location of the two pictures (right, left) was counterbalanced, and the presentation sequence of test questions was randomized. The eight experimental pages were followed by the demographic and language background questions. The whole procedure took around seven minutes.

Predictions

According to previous European-language sound-angularity matching tasks (e.g., D'Onofrio, 2013), we expected participants will show an /i/-pointy, /u/-curvy preference, a 'vowel effect'. If cross-modal perception of pitch-angularity

is guided by sensitivity to Mandarin Chinese tones, we also expected that the Chinese speakers (C & C/E) would show different responses to Tone 1 (T1) and Tone 4 (T4) as T1 has little pitch change (i.e., u1 changes 24Hz; i1 changes 28Hz) while T4 has the most pitch change (see Figure 2.2). On the other hand, the English speakers (E) may show different responses to T1 (high-pitched) and T3 (low-pitched), as English speakers' tone processing mainly focuses on pitch height (Gandour, 1983).

Results

Binomial tests demonstrated that most of the three groups' choices for both /i/ and /u/ were significantly different from the chance level (see Table 2.3). Figure 2.3 shows the percentage of participants who made pointy vs. curvy choices for the different sounds.

Figure 2.3. Angularity selection (%) for the three groups for Study 1A

Table 2.3. Three groups' Binominal Test results of the pointy stimulus compare to the chance level

Chinese-Dominant Group			
Sounds	Observed Proportion	Test Proportion	<i>p</i>
i1	.63	.5	.07
i2	.73	.5	.001
i3	.65	.5	.036
i4	.79	.5	.000
u1	.12	.5	.000
u2	.37	.5	.07
u3	.29	.5	.003
u4	.38	.5	.126

Chinese-English Balanced Bilinguals

Sounds	Observed Proportion	Test Proportion	<i>p</i>
i1	.67	.5	.044
i2	.69	.5	.02
i3	.69	.5	.02
i4	.71	.5	.008
u1	.07	.5	.000
u2	.19	.5	.000
u3	.21	.5	.000
u4	.33	.5	.044

English-speaking Group (no Chinese)

Sounds	Observed Proportion	Test Proportion	<i>p</i>
i1	.8	.5	.000
i2	.68	.5	.028
i3	.78	.5	.000
i4	.88	.5	.000
u1	.29	.5	.012
u2	.17	.5	.000
u3	.17	.5	.000
u4	.22	.5	.000

To investigate the effect of vowel, the effect of tone, the effect of language group and the interactions among them, we ran a generalized linear mixed model test, with vowel, tone, language group, language * vowel, language * tone, vowel * tone, language * vowel * tone as fixed factors and participant as a random factor. Since the outcomes are dichotomous (curvy or pointy), we selected Binary logistic regression that couples a binomial probability distribution with the logit link function (which is the canonical link function for the binomial distribution) (Heck, Thomas, & Tabata, 2012). Since the outcomes for Study 1 (2AFC tasks) were dichotomous, generalized linear model was considered. However, more than one score was observed for each participant,

so participant had to be included as a random effect, with other variables (e.g., tone, shape) as fixed effects. Therefore, I chose the Generalized Linear Mixed Model. The main effect of vowel ($F(1,776)=140.11, p<.001$) and the main effect of tone ($F(3,776)=4.54, p=.004$) were significant (see Figure 2.4 & Table 2.4). The log odds of being curvy for the vowel [u] is increased by 4.123 units ($\beta=4.123, p<.001$) compared with the vowel [i], holding the other effects constant. Overall, all participants, regardless of language backgrounds, made significantly more curvy choices for /u/ stimuli than for /i/ stimuli. For the tone effect, the log odds of being curvy for Tone 1 is increased by 3.178 units ($\beta=3.178, p=.002$) compared with Tone 4, holding the other effects constant. The log odds of being curvy for Tone 2 is increased by 2.603 units ($\beta=2.603, p=.02$) compared with Tone 4, holding the other effects constant. The log odds of being curvy for Tone 3 is increased by 2.234 units ($\beta=2.234, p=.049$) compared with Tone 4, holding the other effects constant. Overall, there was a ‘T1 curvy, T4 pointy, T2 and T3 in between’ response pattern.

Figure 2.4. Estimated Means for significant effects (0=pointy) (Study 1A: overall)

Table 2.4. GLMM Fixed Effects (Study 1A, overall)

	F	df1	df2	Sig.
Vowel	140.108	1	776	.000
Tone	4.538	3	776	.004
Language	1.184	2	776	.307
Language *	1.869	2	776	.155
Vowel				
Language * Tone	1.867	6	776	.084
Vowel * Tone	.959	3	776	.411
Language *	1.09	6	776	.367
Vowel * Tone				

Table 2.5. Descriptive statistics of the pitch of the eight sounds used in this study produced by a female native Mandarin speaker (Hz)

	Onset	Offset	Mean F0	Δ F0 (Max- Min)	Std. Deviation	Variance
u1	281.01	300.98	286.35	24.27	6.2	38.47
u2	236.35	294.17	229.48	81.93	21.74	472.54
u3	209.33	213.31	166.12	47.44	15.38	236.68
u4	308.12	208.79	220.63	140.93	53.73	2886.47
i1	279.7	307.65	296.91	28.47	7.35	54.01
i2	223.07	290.37	229.25	81.13	24.91	620.61
i3	202.16	189.42	181.88	117.64	28.13	791.57
i4	280.52	212.28	224.45	106.94	39.94	1595.36

Given the acoustic differences in pitch between /i/ sounds and /u/ sounds seen in Table 2.5, we also ran separate analyses for the two vowels. For the /i/ sounds, there was a significant main effect of tone ($F(3, 388)=4.05, p=.007$) (see Table 2.6) that did not interact with language groups. The log odds of being pointy for Tone 4 is greater compared with other tones (T1: $\beta=1.962, p=.031$; T2: $\beta=2.912, p=.001$; T3: $\beta=2.476, p=.007$), holding the other effects constant. Overall, significantly more pointy choices compared to curvy choices

were made across the four tones, and this was especially true for T4 (see Figure 2.5).

Figure 2.5. Estimated Means for significant effects (0=pointy) (Study 1A: /i/)

Table 2.6. GLMM Fixed Effects (Study 1A, /i/)

	F	df1	df2	Sig.
Tone	4.052	3	388	.007
Language	1.19	2	388	.305
Language * Tone	1.71	6	388	.117

For the /u/ sounds, there were both significant main effect of tone ($F(3, 388)=4.81, p=.003$) and significant interaction between tone and language group ($F(6, 388)=2.82, p=.011$) (see Table 2.7). The log odds of being curvy

for Tone 1 is increased by 2.1 units ($\beta=2.1$, $p=.006$) in the Chinese-dominant group and increased by 2.436 units ($\beta=2.436$, $p=.008$) in the Chinese-English balanced group compared with the English group, holding the other effects constant. We follow up the interaction between tone and language group further below. As seen in Figure 2.6 (i.e., Tone * Language), overall, the two Chinese groups showed a similar response pattern (i.e., T1 curvy, T4 pointy, T2 & T3 in between), whereas the English group did something different (i.e., T1 and T4, in particular T1, was pointier than T2 and T3). To unpack this interaction effect, we ran pair-wise comparisons between language groups.

Figure 2.6. Estimated Means for significant effects (0=pointy) (Study 1A: /u/)

Table 2.7. GLMM Fixed Effects (Study 1A, /u/)

	F	df1	df2	Sig.
Tone	4.807	3	388	.003
Language	.946	2	388	.393
Language * Tone	2.82	6	388	.011

In the comparison between the CE and the ENG group, the tone effect differed significantly between the CE group and the ENG group ($F(3, 212) = 3.01, p = .031$) (see Table 2.8, Figure 2.6). The log odds of being curvy for Tone 1 is increased by 2.414 units ($\beta = 2.414, p = .01$) in the Chinese-English balanced group compared with the English group, holding the other effects constant. We also conducted pairwise comparisons (Related-Samples McNemar Tests) between tones for /u/ stimuli for each group (see Table 2.9). The pairwise comparisons demonstrated that the CE group made significantly more curvy choices for /u/ stimuli articulated in Tone 1 than in Tone 4 (C/E: $p = .007$), with T2 and T3 between T1 and T4. The English group, however, showed no significant results for any pairwise comparison between tones for the /u/ sound, but the comparison between the pooled T1, T4 and the pooled T2, T3 was significant ($p = .041$). The English-speaking participants made more pointy choices for T1 and T4 than for T2 and T3.

Table 2.8. GLMM Fixed Effects (/u/, CE : ENG)

	F	df1	df2	Sig.
Tone	1.957	3	212	.121
Language	.177	1	212	.674
Language * Tone	3.013	3	212	.031

Table 2.9. Pairwise comparisons between tones

		C/E	C	E
Vowel	Tone	<i>p</i>	<i>p</i>	<i>p</i>
comparison				
u	1&2	.180	.013	.332
u	1&3	.109	.049	.302
u	1&4	.007	.003	.607
u	2&3	1.000	.503	>.999
u	2&4	.180	1.000	.754
u	3&4	.302	.359	.754
u	1&4 vs. 2&3	>.999	.243	.041

In the comparison between C group and E group, the difference in tone effect between the C group and the E group was also significant ($F(3, 272) = 3.78, p = .011$) (see Table 2.10, Figure 2.6). The log odds of being curvy for Tone 1 is increased by 2.04 units ($\beta = 2.04, p = .01$) in the Chinese-dominant group compared with the English group, holding the other effects constant. The results of the Related-Samples McNemar tests showed that, similar to the CE

group, the C group also showed the significant ‘T1-curvy, T4-pointy’ response pattern (C: $p=.003$) (see Table 2.9), with T2 and T3 between them.

Table 2.10. GLMM Fixed Effects (/u/, C: ENG)

	F	df1	df2	Sig.
Tone	1.63	3	272	.183
Language	.754	1	272	.386
Language * Tone	3.784	3	272	.011

When comparing the two Chinese groups, there was a significant main effect of tone ($F(3, 292)=12.28, p<.001$) (see Table 2.11), but this effect did not differ between the two Chinese groups. The log odds of being curvy for Tone 1 is increased by 2.432 units ($\beta=2.432, p<.001$) compared with Tone 4, holding the other effects constant. As seen before, T1 was the curviest, T4 the pointiest, and T2 and T3 between them (see Figure 2.6).

Table 2.11 GLMM Fixed Effects (/u/, C: CE)

	F	df1	df2	Sig.
Tone	12.279	3	292	<.001
Language	1..574	1	292	.211
Language * Tone	.099	3	292	.961

Taking the pairwise comparisons together, it is clear that all Chinese speakers showed a ‘T1-curvy, T4-pointy’ pattern, and both groups differed from English speakers who showed a ‘T1 & T4 pointy, T2 & T3 curvy’ pattern.

Discussion

Consistent with the results of previous studies (Köhler, 1929; Ramachandran & Hubbard, 2001; D’Onofrio, 2013), ‘*u*-curvy, *i*-pointy’ angularity-vowel cross-modal correspondences have been replicated by our online 2AFC survey in a variety of populations, suggesting that these cross-modal correspondences might be universal, at least within humans.

In addition to vowel features, Chinese speakers (both C Group and C/E Group) made much more curvy choices for /u/ stimuli articulated in Tone 1 than in Tone 4, which is consistent with our predictions that Chinese speakers’ sensitivity to Mandarin tones would influence their audiovisual cross-modal perception. That is, they make the most curvy choices when hearing the least changing T1 while the most pointy choices when hearing the most changing T4 (see Table 2.5, Figure 2.2), suggesting that pitch steadiness of lexical tone enhances the *u*-curvy vowel-angularity mapping, while pitch change weakens this vowel-shape mapping.

By contrast, English-speaking participants made fewer curvy choices for Tone 1 and Tone 4 compared to for Tone 2 and Tone 3 for. The similarity between Tone 1 and Tone 4 is that they are both high in pitch on average and have higher pitch onsets compared to Tone 2 and Tone 3 (see Table 2.5, Figure 2.2). In other words, instead of pitch change, the English speakers tend to pay more attention to pitch height. The higher the pitch, the pointier the shape of its association. Taken together, the finding of the English-speaking participants is

consistent with the ‘high-pointy, low-curvy’ pattern observed in the existing literature on the relationship between pure tones and shapes. On the other hand, the finding of the Chinese-speaking participants (C; C/E) is different from the existing findings. Chinese speakers match Mandarin tones with shapes depending on pitch change within each tone.

To sum up, Chinese speakers (both C Group and C/E Group) who are more sensitive to pitch change, tend to match shapes differing in angularity with sounds in terms of pitch change. That is, the greater the pitch change within a Mandarin tone, the pointier the shape of its association. This happened only to the sounds of /u/. /i/, however, was always matched with the pointy shape regardless of being articulated in which tone across different language groups. This suggests that the vowel effect might outperform lexical tone effect in this kind of cross-modal correspondence.

The English speakers with no experience of Chinese tones (E Group), on the other hand, tend to pay more attention to pitch onset than pitch change. They tend to link sounds with higher pitch onset to pointer shapes and sounds with lower pitch onset to curvier shapes. This finding is quite similar to the ‘high-pitched sound-pointy, low-pitched sound-curvy’ pattern reported in the pitch-angularity audiovisual cross-modal perception literature (Marks, 1987; O’Boyle & Tarte, 1980; Walker et al., 2010), except that changing lexical tones were used in the present study.

In the present study, participants were presented with a pair of shapes with different angularity and a single sound articulated in a certain Mandarin tone and asked to match the sound with one of the shapes. However, hearing the sounds in isolation might be too difficult for the English speakers who are less sensitive to the changing tones. For this reason, we developed a second study designed to make the differences between different pitch contours more obvious even for non-Tone language participants, by presenting two sounds very close to one another (so that they can actually hear the differences) to see whether this change influences the way they respond, or whether pitch height is still the main driver of the effect.

2.3 Study 1B: Two sounds with one shape

Participants

There were 104 participants (68 females), aged 18 to 57 years, participated in this online survey (31 participants in Chinese dominant bilingual group (C); 24 participants in Chinese-English balanced bilingual group (C/E) and 49 participants in English speakers with no Chinese group (E)). The classification criteria for the language groups were identical to that of Study 1b.

Stimuli

The same stimuli were used in Study 1B. In study 1B, each minimal pair for Mandarin tone contrasts for each vowel (i.e., i1-i2; i1-i3; i1-i4; i2-i3; i2-i4; i3-

i4; u1-u2; u1-u3; u1-u4; u2-u3; u2-u4; u3-u4) was paired with either the curvy or the pointy shape. The same tone contrasts with different shapes (e.g. i1-i3 with the pointy shape; i1-i3 with the curvy shape) were counterbalanced across participants. Given our interest in whether pitch change or pitch height is the main driver of tone-shape matching, we restrict our analysis to the T1/T3 and the T1/T4 pairs.

Procedure

In each question, participants were presented with one picture of a shape (curvy or pointy) and a button which triggered the audio file including a minimal pair of sounds to play which were separated by a brief silence of 500ms duration. The participants were asked which of the two sounds (the first or the second) went better with the shape. The remaining procedures were identical to that of Study 1A.

Results

Given the findings that in Study 1A pitch change is important for Chinese speakers and pitch height is important for English speakers in their tone-shape matchings, as well as the fact that T1 and T3 contrast the most in terms of pitch height and T1 and T4 differ the most in pitch change (see Table 2.5 and Figure 2.2), our interest here is whether pitch-change or pitch-height is the main driver of tone-shape matchings for different groups of participants. Therefore, here,

we focus on the results of the T1-T3 pair and those of T1-T4 pair, which are maximally contrastive for height, and change, respectively.

For the T1-T3 contrast, we first ran a generalized linear mixed model test, with shape (curvy, pointy), language, shape * language interaction as fixed factors and participant as a random factor. The main effect of shape ($F(1, 202) = 20.469, p < .001$) and the interaction between shape and language ($F(2, 202) = 8.027, p < .001$) were significant (see Table 2.12). The log odds of being Tone 1 for the curvy shape is decreased by 2.81 units ($\beta = -2.81, p < .001$) compared with the pointy shape, holding the other effects constant. The log odds of being Tone 1 for the curvy shape is increased by 2.944 units ($\beta = 2.944, p < .001$) in the Chinese-dominant group compared with the English group, holding the other effects constant. Overall, regardless of whether the sound is /i/ or /u/, three subject groups showed significantly different mapping patterns between the shapes and the tones (see Figure 2.7). From Figure 2.7, it seemed like that both the CE group and the E group showed a significant ‘T3-curvy, T1-pointy’ pattern, whereas the C group did not show a clear mapping pattern. To unpack this interaction effect, further pairwise comparisons have been done.

Figure 2.7. Estimated Means for significant effects (0=T3) (Study 1B, T1-T3: overall)

Table 2.12. GLMM Fixed Effects (Study 1B, T1-T3, overall)

	F	df1	df2	Sig.
Shape	20.469	1	202	<.001
Language	2.753	2	202	.066
Language *	8.027	2	202	<.001
Shape				

When comparing the CE group with the E group, there was only a significant main effect of shape ($F(1, 142) = 30.046, p < .001$) (see Table 2.13). The log odds of being Tone 1 for the curvy shape is decreased by 2.81 units ($\beta = -2.81, p < .001$) compared with the pointy shape, holding the other effects constant. Both groups showed a significant ‘T1-pointy, T3-curvy’ mapping

pattern between shapes and tones without a significant group difference (see Figure 2.7).

Table 2.13. GLMM Fixed Effects (Study 1B, T1-T3, CE vs. E)

	F	df1	df2	Sig.
Shape	30.046	1	142	<.001
Language	3.536	1	142	.062
Language * Shape	1.289	1	142	.258

When comparing the C group and the E group, the main effect of Shape ($F(1, 156)=12.936, p<.001$), the main effect of Language ($F(1, 156)=4.385, p<.038$), and the interaction between them ($F(1, 156)=15.659, p<.001$) were all significant (see Table 2.14). The log odds of being Tone 1 for the curvy shape is increased by 2.944 units ($\beta=2.944, p<.001$) in the Chinese-dominant group compared with the English group, holding the other effects constant. To unpack this interaction effect, we also conducted pairwise comparisons (Related-Samples McNemar Tests) between shapes for each group. The results demonstrated that the E group showed a significant ‘T1-pointy, T3-curvy’ mapping pattern ($p<.001$), on the other hand, the C group did not show a significant mapping pattern between the tones and the shapes for the T1-T3 contrast. And as seen above, the difference in the tone-shape mappings between the two groups was significant ($F(1, 156)=15.659, p<.001$).

Table 2.14. GLMM Fixed Effects (Study 1B, T1-T3, C vs. E)

	F	df1	df2	Sig.
Shape	12.936	1	156	<.001
Language	4.385	1	156	.038
Language *	15.659	1	156	<.001
Shape				

The comparison between the C group and the CE group showed a significant main effect of Shape ($F(1, 106)=4.04, p=.047$) and a significant interaction between language and shape ($F(1, 106)=5.405, p=.022$) (see Table 2.15). The log odds of being Tone 1 for the curvy shape is increased by 1.98 units ($\beta=1.98, p=.022$) in the Chinese-dominant group compared with the Chinese-English balanced group, holding the other effects constant. For this interaction effect, we also conducted pairwise comparisons (Related-Samples McNemar Tests) between shapes for each group. The results demonstrated that like the E group, the CE group also showed a significant ‘T1-pointy, T3-curvy’ mapping pattern ($p=.013$), which was similar to what the E group did ($F(1, 142)=1.289, p=.258$), but significantly different from what the C group did ($F(1, 106)=5.405, p=.022$).

Table 2.15. GLMM Fixed Effects (Study 1B, T1-T3, C vs. CE)

	F	df1	df2	Sig.
Shape	4.04	1	106	.047
Language	.002	1	106	.963
Language *	5.405	1	106	.022
Shape				

For the T1-T4 contrast, we also first ran a generalized linear mixed model test, with shape (curvy, pointy), language, shape * language interaction as fixed factors and participant as a random factor. The main effect of shape ($F(1, 202)=7.829, p=.006$) and the interaction between shape and language ($F(1, 202)=7.353, p=.001$) were significant (see Table 2.16). The log odds of being Tone 1 for the curvy shape is increased by 2.559 units ($\beta=2.559, p<.001$) in the Chinese-dominant group and increased by 1.919 units ($\beta=1.919, p=.015$) in the Chinese-English balanced group compared with the English group, holding the other effects constant. Again, overall, regardless of whether the sound is /i/ or /u/, the difference in the mapping patterns between the shapes and the tones was significant across subject groups (see Figure 2.8). From Figure 2.8, it seemed like that the two Chinese groups (i.e., C & CE) showed a significant ‘T1-curvy, T4-pointy’ pattern, whereas the E group showed an opposite mapping pattern. We did pairwise comparisons to explore this interaction effect.

Figure 2.8. Estimated Means for significant effects (0=T4) (Study 1B, T1-T4: overall)

Table 2.16. GLMM Fixed Effects (Study 1B, T1-T4, overall)

	F	df1	df2	Sig.
Shape	7.829	1	202	.006
Language	.639	2	202	.529
Language *	7.353	2	202	.001
Shape				

The interaction between language and shape was significant in the comparison between the CE group and the E group ($F(1, 142)=6.086, p=.015$) (see Table 2.17). The log odds of being Tone 1 for the curvy shape is increased by 1.919 units ($\beta=1.919, p=.015$) in the Chinese-English balanced group

compared with the English group, holding the other effects constant. We also conducted pairwise comparisons (Related-Samples McNemar Tests) between shapes for each group. The results demonstrated that the CE group showed a significant ‘T1-curvy, T4-pointy’ mapping pattern ($p=.039$), and the E group showed a ‘T1-pointy, T4-curvy’ mapping pattern, but not significant ($p=.248$) (see Figure 2.8). While the difference in the tone-shape mapping pattern between the two groups was significant ($F(1, 142)=6.086, p=.015$).

Table 2.17. GLMM Fixed Effects (Study 1B, T1-T4, CE vs. E)

	F	df1	df2	Sig.
Shape	.931	1	142	.336
Language	1.144	1	142	.287
Language * Shape	6.086	1	142	.015

There was also a significant interaction effect between language and shape in the comparison between the C group and the E group ($F(1, 156)=12.723, p<.001$) (see Table 2.18). The log odds of being Tone 1 for the curvy shape is increased by 2.559 units ($\beta=2.559, p<.001$) in the Chinese-dominant group compared with the English group, holding the other effects constant. The results of the Related-Samples McNemar Test demonstrated that like the CE group, the C group also showed a significant ‘T1-curvy, T4-pointy’ mapping pattern

($p=.001$), which was significantly different from the E group's weak 'T1-pointy, T4-curvy' mapping pattern ($F(1, 156)=12.723, p<.001$) (see Figure 2.8).

Table 2.18. GLMM Fixed Effects (Study 1B, T1-T4, C vs. E)

	F	df1	df2	Sig.
Shape	3.753	1	156	.055
Language	<.001	1	156	.986
Language *	12.723	1	156	<.001
Shape				

Lastly, the comparison between the two Chinese groups (i.e., C & CE) showed a significant main effect of shape ($F(1, 106)=14.042, p<.001$), which did not differ significantly between the two groups (see Table 2.19). The log odds of being Tone 1 for the curvy shape is increased by 1.335 units ($\beta=1.335, p=.046$) compared with the pointy shape, holding the other effects constant. Overall, the Chinese speakers shared a 'T1-curvy, T4-pointy' tone-shape mapping pattern (see Figure 2.8).

Table 2.19. GLMM Fixed Effects (Study 1B, T1-T4, C vs. CE)

	F	df1	df2	Sig.
Shape	14.042	1	106	<.001
Language	<.001	1	106	.341
Language *	12.723	1	106	.471
Shape				

Discussion

Overall, when the vowel was controlled, for the T1-T3 contrast, there was a significant difference in tone-shape mappings across subject groups. Both the English-speaking participants (E) and the Chinese-English balanced participants (CE) showed no group difference in their significant ‘T1-pointy, T3-curvy’ tone-shape mapping pattern. However, the Chinese-dominant participants (C) did not show a clear mapping pattern and the C group differed significantly in this tone-mapping correspondence from the other groups. As seen in Figure 2.2 and Table 2.5, in terms of average pitch, Tone 1 is the highest and Tone 3 is the lowest. Therefore, T1-T3 is the best high-low contrast in the four Mandarin tones, which could probably explain why the CE group, like the C group showed the pitch-change-driven ‘T1-curvy, T4-pointy’ tone-shape mapping pattern in Study 1A, while they show the pitch-height-driven pattern here when presented with the T1-T3 contrast. First, the best Mandarin high-low contrast (i.e., T1-T3) might trigger the Chinese-English balanced bilinguals’ English-like pitch-height-driven tone perception and tone-shape crossmodal correspondence. In addition, the task, itself, may also encourage the English speakers’ fairly strong sensitivity to pitch height as well as the CE balanced bilinguals’ by controlling the vowels to make the tone contrast more salient. Furthermore, all tasks were carried out in English, which may have contributed to the balanced bilinguals’ more English-like pattern here. For the Chinese-dominant bilinguals (C), their weaker English compared to the CE

group (see Table 2.1), and in turn, their weaker sensitivity to pitch height but stronger sensitivity to pitch change might speak for the unclear pattern of their response to the better pitch-height-driven but worse pitch-change-driven tone contrast between T1 and T3. Furthermore, the high (i.e., T1)-pointy, low (i.e., T3)-curvy pattern observed in the CE group and the E group is consistent with the findings in the literature on the relationship between visual angularity and pitch height of pure tones (Marks, 1987; O'Boyle & Tarte, 1980; Walker et al., 2010).

The Pitch-Change Hypothesis set up in Study 1A can help explain the Chinese participants' (i.e., C & CE) results of the T1-T4 contrast. In Study 1A, overall, the two groups showed a significant 'T1-curvy, T4-pointy' mapping pattern without significant group difference between the two Chinese groups but with a significant difference from that of the E group. Again, as seen in Figure 2.2 and Table 2.5, in terms of pitch change, Tone 4 changes the most and Tone 1 changes the least. Therefore, T1-T4 is the best pitch-change-driven contrast among the Mandarin tones. In addition, pitch change is very important in Chinese speakers' pitch perception (see Jongman, et al., 2006; Singh, et al., 2014 for more discussion on this point). This might be the reason why the two Chinese groups consistently matched the more changing Tone 4 with the pointy shape and the less changing T1 with the curvy shape. The English-speaking participants (E) showed a different tone-shape matching pattern from the Chinese groups, that is, a 'T1-pointy, T4-curvy' pattern. However, this English mapping pattern was not significant. Again, English speakers are sensitive to

pitch height, because it is important in their pitch perception (Gandour, 1983). However, T1-T4 is a better pitch-change-driven contrast but a worse one for pitch height. That might be the reason why the English speakers did not do well for the ‘high (T1)-pointy, low (T4)-curvy’ pattern here (T1 was on average higher than T4, see Figure 2.2 and Table 2.5).

To sum up, in terms of the relationship between visual angularity and Mandarin Tones, the English-speaking participants (E) showed a ‘high-pointy, low-curvy’ mapping pattern, while the Chinese-dominant participants (C) showed a pitch-change pattern. That is, the greater the pitch-change within a Mandarin tone, the pointier the shape of its association. The Chinese-English balanced participants (CE) however, seemed to swing between the pitch-height mapping pattern and the pitch-change pattern depending on task types and test tonal contrasts, showing perceptual flexibility.

General Discussion

Our online 2AFC survey replicated the ‘*u*-curvy, *i*-pointy’ visual angularity-vowel cross-modal correspondences, lending support for the universality of this kind of correspondence. The English-speaking participants (E) turned out to match visual angularity with Mandarin tones according to pitch height. That is, high-pitched sounds were matched to pointy shapes, while low-pitched sounds with curvy shapes. The Chinese dominant bilinguals (C), on the other hand, matched visual angularity with Mandarin tones in terms of pitch change. That is,

the greater the pitch-change within a Mandarin tone, the pointier the shape of its association. The Chinese-English balanced bilinguals (C/E) seemed to swing between the two strategies depending on task types and tonal contrasts. Moreover, the vowel effect seemed to outdo lexical tone effect on this kind of cross-modal correspondences across all the subject groups. Taken together, language background seems to tune multisensory perception between visual angularity and lexical tones.

Chapter 3: IAT of Shapes and Vowels with Tones

3.1 Introduction

In the previous chapter, the ‘*u*-curvy, *i*-pointy’ effect was replicated using 2AFC tasks. Moreover, different participant groups showed different preference patterns for tone-shape mappings. In this chapter, instead of overtly asking about people’s preference of sound-shape mappings, we used Implicit Association Test (IAT) tasks to investigate people’s covert cognitive biases for sound-shape mappings. The IAT measures how strong the associations are between items (e.g., white people, cats) and evaluations (e.g., good, cute). The IAT was introduced in scientific research by Greenwald, McGhee and Schwartz in 1998 and has been widely used in social psychology research (e.g. implicit racism research). The IAT has also been used in cross-modal correspondence investigations. Using the IAT, Parise and Spence (2012) have replicated several classical findings in the sound symbolism literature (i.e., takete/maluma; mil/mal) together with a finding of the mapping between visual size and auditory pitch in the literature. Moreover, using the IAT, they first found cross-modal correspondence between visual size of angles and auditory pitch as well as correspondence between the sharpness of visual shapes and the waveform of auditory stimuli. In this chapter, to avoid ‘strategic’ responses, we investigated the Pitch-Change Hypothesis (e.g., T1-curvy, T4-pointy) and the Pitch-Height Hypothesis (T1-pointy, T3-curvy) using the IAT in Chinese dominant bilinguals (C), balanced bilinguals (C/E) and English speakers (E). We

examined the effect of lexical tone-visual angularity congruence on reaction times by manipulating mappings between stimuli and response keys. We also manipulated attention to different acoustic features of sounds to see how attention modulates participants' tone-angularity cross-modal correspondences.

3.2 Study 2A: IAT of Shapes and Vowels in Tone 1 & 4 – Vowel Attended

Participants

We recruited participants for the IAT tasks mainly using Nanyang Technological University Division of Psychology's Research Participation (RP) pool. The Chinese-English balanced participants (C/E) were all from our RP pool and their participation was for course credit. While the Chinese dominant bilinguals (C) and the English-speaking participants (E) were partially from the RP pool and partially recruited through printed advertisement on notice boards at NTU. The participants who were not for credits received S\$8 for their participation. The present study was funded by Prof Suzy Styles' Nanyang Assistant Professorship SUG. The classification criteria for participant groups were identical to Study 1. The participant recruitment channels and the participant group classification remained the same for all the subsequent studies. In the present study, the sample size for C/E group was $n = 20$, while due to the recruitment difficulty (i.e., scarcity of monolingual English speakers and Chinese dominant bilinguals who grew up outside Singapore), the sample size

for C group and E group was $n = 10$ respectively. For a summary of age and gender for all groups of participants, see Appendix 3.1.

Apparatus

The presentation of the stimuli and the response collection were controlled by the desktop computers running E-Prime 2.0 (SP2) (2.0.10.356) located in the psychological laboratories in NTU HSS. Each desktop used in the present study has a 19" LCD monitor with a resolution of 1280×1024 pixels.

Stimuli

We used the same visual stimuli as those used in Study 1. For the auditory stimuli, we used the sound files of /i1/, /i4/, /u1/ and /u4/ used in Study 1A.

Procedure

When participants came to the laboratory, they first received and signed the Information & Consent sheet approved by the IRB of Nanyang Technological University. Then each participant was seated in front of a desk and put on a pair of headphones. Prior to the experimental trials, they were instructed to do some practice trials to familiarize themselves with the experimental paradigm. The stimuli used in the practice trials were different from those used in experimental trials, but the paradigm was identical. At the beginning of the practice phase, the participants were presented with the instructions (see Figure 3.1) which informed them that when seeing a green square or hearing a violin sound they

were supposed to press the “A” key, while when seeing a blue square or hearing a trumpet sound they should press the “L” key.

There was a green square, a blue square, but two violin sounds (one C4, one E4) and two trumpet sounds (one C4, one E4) used. This meant that one feature of the audio (instrument) was attended due to the button-press instruction, while another feature of the audio (pitch) was not attended. As the practice phase was to get participants familiar with idea of mappings between stimuli and responses, the results of practice trials were not analyzed.

Figure 3.1. Illustration of stimulus-response key assignment for practice trials

In the experimental phase, the participants were shown the overall instruction first, followed by the two pictures (one curvy, one pointy) that they would see. Then participants were informed that the sounds they would hear are

the “ee” vowel in “feet” and the “oo” vowel in “shoes”. They were also instructed to keep their eyes focused on the “+” in the center of the screen and respond as quickly and as accurately as possible by pressing the appropriate response key. We replicated the IAT procedures used in Parise and Spence (2012). Each trial began with a black cross in the center of the screen presented for a randomized duration of 500-600 ms. After the removal of the fixation, there was a random ISI (inter-stimulus interval) of 500-600 ms, which was followed by a target stimulus, either visual or auditory. Each visual stimulus was presented for 500 ms. Each auditory stimulus lasted until the sound file ended, which did not exceed 500 ms. Each auditory stimulus was a single vowel (/i/ or /u/) articulated in Tone 1 or Tone 4 and was presented only once on each trial. Incorrect responses and no responses within 10s from the presentation of the stimulus were followed by feedback in the form of a red cross presented in the center of the screen for 500 ms (see Figure 3.2.)

(A)

(B)

Figure 3.2. The timeline for experimental trials in Study 2A, showing an example of trials in a congruent block: (A) an example of a visual trial and (B) an example of an auditory trial.

At the beginning of each experimental block, participants were presented with the instruction of the mappings between the stimuli and the response keys for the upcoming block (see Figure 3.3). Within each experimental block, one shape and one vowel were assigned to one response key (either “A” or “L”) and the other vowel and the other shape were assigned to the other key. The instruction remained on the screen until the participants started a new block by pressing the space bar. The manipulation of the mapping between the stimuli and response keys for the experimental blocks resulted in four different pairings. Two of them were considered as angularity-vowel congruent blocks where /u/ and the curvy shape were assigned to an identical response key (either “A” or “L”) and /i/ and the pointy shape were assigned to the other response key. While in the two angularity-vowel incongruent blocks, /u/ and the pointy shape were mapped onto an identical response key and /i/ and the curvy shape were mapped onto the other key. Note that each vowel was articulated in both Tone 1 and Tone 4, although in the present study (Study 2A), the participants were not instructed to pay attention to the tones of the vowels. Therefore, besides vowel-angularity congruence, there was tone-angularity congruence as well which will be addressed in detail in the Result section.

Each block (i.e. each type of pairing) consisted of eight items of the curvy shape, eight items of the pointy shape, four /i1/s, four /i4/s, four /u1/s and four /u4/s. There were four block types (two congruent, two incongruent) and each block type repeated four times, ending up with a total of 512 trials. Participants

could take breaks during the intervals between blocks. The whole process took about 40 minutes. Reaction times (RT) and Accuracy were collected.

Figure 3.3. Examples of instructions used in the experimental blocks in Study 2A. Left: An example of the instruction for an angularity-vowel congruent block. Right: An example of the instruction for an angularity-vowel incongruent block.

Data handling and analysis

Accuracy was computed for each subject group. The distribution of raw reaction times were not normally distributed, which is common in RT measures. The distribution was skewed towards the left with a tail to the right. Reaction times were transformed into inverse RT ($iRT=1/RT$) to approximate normal distribution. To trim extreme data in the current datasets, responses greater than three standard deviations above or below the mean for each participant were excluded. After the application of the cutoffs, distributions were checked again to ensure symmetry. The distributions were roughly symmetrical, and thus the mean would be an appropriate measure of central tendency, allowing parametric statistics to be used. All analyses were conducted on mean iRT , while back-transformed RT ($RT=1/iRT$) was used for

visual presentation and reporting means. Primary analyses were conducted in SPSS (Version 23), with supplementary pairwise Power analyses conducted in G*Power (Version 3.1, Faul, Erdfelder, Lang, & Buchner, 2007). The same procedure was used for all of the subsequent experiments.

Results

The accuracy of the three subject groups was 96.6% (C), 93.1(C/E) and 94.2% (E) respectively. Only auditory trials were analyzed for Study 2. In all the studies of Study 2 (i.e., 2A, 2B, 2C & 2D), auditory stimuli had two dimensions (vowel and tone) while visual stimuli had one dimension (angularity). Visual stimuli could only be analyzed in terms of the block-level congruence, and are thus excluded from the analysis. Given the low number of errors, incorrect responses were excluded from the analysis. Vowel-angularity congruence effect, tone-angularity congruence effect and the interaction between them were of interest. The inverse RTs of the auditory trials were submitted to a repeated-measure analysis of variance (ANOVA) with the within-subjects factors of vowel-angularity congruence (congruent versus incongruent) and tone-angularity congruence (congruent versus incongruent) (see Table 3.1). Vowels were the attended dimension in the present study, and vowel-angularity is therefore the attended congruence dimension. The ‘hidden’ congruence dimension was tone-angularity congruence, and the tone-angularity congruence used here was the Chinese speakers’ mapping pattern observed in Study 1 (i.e., The greater the pitch change within a Mandarin tone, the pointier

the shape of its association. That is, ‘T1-curvy, T4-pointy’ is congruent and ‘T1-pointy, T4-curvy’ is incongruent.). We examined this Chinese lexical tone-angularity mapping pattern in a diverse population using IAT task. Figure 3.3 displays the back-transformed RTs (back-transformed $RT=1/iRT$) shown separately for the three experimental groups, and the different congruence conditions.

Table 3.1. Different conditions of Vowel-angularity congruence and Tone-angularity congruence in Study 2A

	Congruent condition	Incongruent condition
Vowel-angularity (Attended)	/u/-curvy; /i/-pointy	/u/-pointy; /i/-curvy
Tone-angularity (Unattended)	T1-curvy; T4-pointy	T1-pointy; T4-curvy

Figure 3.4. Bar charts for different subject groups' back-transformed RTs in Study 2A (IAT of Shapes and Vowels in Tone 1 & 4: Vowel Attended) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

Chinese-Dominant Group

The results of the Chinese dominant group (C) showed a significant main effect of vowel-angularity congruence ($F(1,9)=12.29, p=.007, \eta_p^2=.58$), indicating that vowel-angularity congruent trials ($M=691, SD=105$) were processed significantly faster than vowel-angularity incongruent trials ($M=770,$

SD=114), regardless of tone-angularity congruence. No other effects or interactions were significant.

Table 3.2. ANOVA results of iRT (Chinese-dominant bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	12.29	.007	.577	.876
Tone Congruence	1	.15	.711	.016	.064
Interaction	1	3.74	.085	.294	.409

Chinese-English Balanced Bilinguals

The balanced bilingual group (C/E) also showed a significant main effect of vowel-angularity congruence only ($F(1,19)=26.51$, $p<.001$, $\eta_p^2=.58$). The results indicated that vowel-angularity congruent trials ($M=642$, $SD=103$) were significantly faster than vowel-angularity incongruent trials ($M=737$, $SD=119$).

Table 3.3. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	26.51	<.001	.583	.998
Tone Congruence	1	.02	.888	.001	.052
Interaction	1	.012	.914	.001	.051

English-speaking Group (non-Chinese)

English-speaking participants (E) also showed a significant main effect of vowel-angularity congruence only ($F(1,9)=11.79$, $p=.007$, $\eta_p^2=.57$). Vowel-angularity congruent trials ($M=684$, $SD=114$) were significantly faster than incongruent ones ($M=740$, $SD=103$).

Table 3.4. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	11.79	.007	.567	.862
Congruence					
Tone	1	1.51	.251	.143	.196
Congruence					
Interaction	1	.13	.729	.014	.062

A comparison among the three groups

To investigate the difference in the effects across language groups, we ran a 3-way repeated measures ANOVA, with vowel-angularity congruence (congruent, incongruent) and tone-angularity congruence (congruent, incongruent) as within-subjects factors and language group (C, C/E, E) as a between-subjects factor. To keep the equal sample size across language groups ($n=10/\text{group}$), we selected the first 10 participants for the C/E group. The only significant effect was the main effect of vowel-angularity congruence ($F(1,27)=27.92$, $p<.001$, $\eta_p^2=.508$). Vowel-angularity congruent trials ($M=684$, $SD=103$) were significantly faster than incongruent ones ($M=748$, $SD=117$) regardless of tone-angularity congruence effect, and this main effect did not differ between the language groups.

Table 3.5. ANOVA results of iRT (overall Study 2A)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	27.92	<.001	.508	>.999
VC * Language	2	.45	.642	.032	.116
Tone Congruence	1	.1	.751	.004	.061
TC * Language	2	.5	.613	.036	.123
VC * TC	1	.03	.862	.001	.053
VC*TC*Language	2	1.97	.159	.127	.372
Language Group	2	.23	.795	.017	.083

Discussion

When vowels were paid attention to, all participants of the three subject groups showed a strong main effect of vowel-angularity (i.e., /u/-curvy; /i/-pointy) regardless of whether tones of the vowels were cross-modally congruent with the angularity of the shapes or not. These results replicated the findings of vowel-angularity mappings in the literature (e.g., Köhler 1929:1947; Ramachandran and Hubbard, 2001; Maurer et al., 2006; Nielsen and Rendall, 2011; 2013; D’Onofrio, 2014). No significant tone effects were observed. This finding (and the power >.999) validates the method for the main effect of vowel-angularity congruence with Ns as small as 10.

3.3 Study 2B: IAT of Shapes and Vowels in Tone 1 & 4 – Tone Attended

In order to see whether attention to a target feature of auditory stimuli makes difference in cross-modal correspondence between sounds and shapes, tone was attended in the present study. We used the same stimuli (i.e., two shapes: one curvy, one pointy; four sounds: i1, i4, u1, u4) and procedures as those in Study 2A, except that in each experimental block, instead of vowels, one shape and one tone were assigned to one response key (either “A” or “L”) and the other tone and the other shape were assigned to the other key. The instruction of the mapping between the stimuli and the response keys for the upcoming block appeared at the beginning of each block. In the present study, the manipulation of the tone-angularity congruence was based on the Chinese speakers’ lexical tone-visual angularity mapping pattern. Therefore, in the two angularity-tone congruent blocks, T1 (the high-level tone) and the curvy shape were assigned to an identical response key (either “A” or “L”) and T4 (the falling tone) and the pointy shape were assigned to the other response key. While in the two angularity-tone incongruent blocks, T1 and the pointy shape were mapped onto an identical response key and T4 and the curvy shape were mapped onto the other key (see Figure 3.5). By contrast to Study 2A, the vowels were therefore unattended here. Ten participants fell into the C group, twenty fell into the C/E group and ten in the E group. Since non-Chinese speakers (E) have no knowledge of Mandarin tones, Tone 1 was described as “High sound” and Tone 4 was described as “Falling sound” in the instructions

for the E Group. The non-Chinese speaking participants were also instructed to listen to the stimuli as much as they wanted, until they were sure that they could correctly identify which was the High sound and which was the Falling sound.

Figure 3.5. Examples of instructions used in the experimental blocks in Study 2B for the two Chinese groups. Left: An example of the instruction for an angularity-tone congruent block. Right: An example of the instruction for an angularity-tone incongruent block.

Results

The accuracy of the three subject groups were 94% (C), 93.1% (C/E) and 89% (E) respectively. Data handling was conducted as in study 2A and data analysis was performed using the same procedures as in Study 2A, except that instead of vowel-angularity congruence, tone-angularity congruence was the attended dimension in the present study. Figure 3.6 displays the back-transformed RTs shown separately for the three experimental groups, and the different congruence conditions.

Figure 3.6. Bar charts for different subject groups' back-transformed RTs in Study 2B (IAT of Shapes and Vowels in Tone 1 & 4: Tone Attended) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

Chinese-Dominant Group

No main effect of vowel-angularity congruence or main effect of tone-angularity congruence was detected (see Table 3.6). However, a significant interaction between tone effect and vowel effect was significant ($F(1,9)=15.18$, $p=.004$, $\eta_p^2=.63$). To unpack the pattern of interactions, a paired-samples t-test was conducted to compare the inverse RTs between tone-angularity congruent

condition and incongruent condition when vowel and visual angularity were incongruent. There was a significant difference in inverse RTs between tone-angularity congruent condition ($M=776$, $SD=149$) and incongruent ($M=862$, $SD=160$) condition, $t(9) = 3.932$, $p=.003$. Tone-angularity congruent trials were significantly faster than incongruent ones. A paired-samples t-test was also performed to compare the inverse RTs between tone-angularity congruent condition and incongruent condition when vowel and visual angularity were congruent. There was no significant difference in inverse RTs between tone-angularity congruent condition ($M=818$, $SD=178$) and incongruent condition ($M=818$, $SD=174$) ($t(9) = .154$, $p=.881$).

Table 3.6. ANOVA results of iRT (Chinese-dominant bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	.23	.642	.025	.072
Congruence					
Tone	1	3.98	.077	.307	.43
Congruence					
Interaction	1	15.18	.004	.628	.933

Chinese-English Balanced Bilinguals

No main effect of vowel-angularity congruence or main effect of tone-angularity congruence was detected (see Table 3.7). However, the interaction between tone effect and vowel effect was significant ($F(1,19)=9.53$, $p=.006$, $\eta_p^2=.33$). As above, to unpack the pattern of interactions, paired-samples t-tests were conducted. Similarly, there was a significant difference in inverse RTs between tone-angularity congruent condition ($M=764$, $SD=117$) and incongruent condition ($M=845$, $SD=166$) ($t(19)=3.237$, $p=.004$), when vowel and visual angularity were incongruent. Tone-angularity congruent trials were significantly faster than incongruent ones. While there was no significant difference in inverse RTs between tone-angularity congruent condition ($M=816$, $SD=157$) and incongruent condition ($M=813$, $SD=150$) ($t(19)=-.106$, $p=.917$), when vowel and visual angularity were congruent. The results suggested that when vowel and visual angularity were incongruent, both of the two Chinese groups responded to tone-angularity congruent trials much faster than tone-angularity incongruent ones.

Table 3.7. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	1.92	.182	.092	.261
Congruence					
Tone	1	3.76	.068	.165	.452
Congruence					
Interaction	1	9.53	.006	.334	.833

English-speaking Group (non-Chinese)

There was a significant main effect of tone-angularity congruence ($F(1,9)=21.28, p=.001, \eta_p^2=.7$), suggesting that tone incongruent trials ($M=978, SD=133$) were significantly faster than tone congruent ones ($M=1107, SD=217$). The interaction between vowel congruence and tone congruence was also significant ($F(1,9)=5.86, p=.039, \eta_p^2=.39$). The results of a paired-samples t-test showed that there was a significant difference in inverse RTs between tone-angularity congruent condition ($M=1123, SD=243$) and incongruent condition ($M=934, SD=122$) when vowel and visual angularity were congruent ($t(9) = -4.84, p=.001$). Tone-angularity incongruent trials were significantly faster than congruent ones. A paired-samples t-test was also performed to compare the inverse RTs between the tone-angularity congruent and incongruent conditions

when vowel and visual angularity were incongruent. There was no significant difference in inverse RTs between tone-angularity congruent ($M=1090$, $SD=191$) and incongruent ($M=1022$, $SD=143$) conditions ($t(9) = -1.67$, $p=.129$).

Table 3.8. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	.03	.859	.03	.05
Congruence					
Tone	1	21.28	.001	.7	.98
Congruence					
Interaction	1	5.86	.039	.39	.58

A comparison among the three groups

We also ran a 3-way repeated measures ANOVA, with vowel-angularity congruence (congruent, incongruent) and tone-angularity congruence (congruent, incongruent) as within-subjects factors and language group (C, C/E, E) as a between-subjects factor, to test whether the simple effects differ between the levels of language group. To keep the equal sample size across language groups ($n=10/\text{group}$), we selected the first 10 participants for the C/E group. Along with the previously observed interaction between vowel congruence and tone congruence, this analysis revealed a main effect of

language group (such that English speakers were slower overall). There was also a significant interaction between tone-angularity congruence effect and language group ($F(2,27)=14.81, p<.001, \eta_p^2=.52$). To unpack this interaction, we first computed the difference between the averaged inverse RTs in tone congruent trials and those in tone incongruent trials (i.e., congruent – incongruent) and then we conducted pair-wise independent-samples T Tests between the language groups. The results suggested that the non-Chinese group ($M=123, SD=87$) had a difference in the opposite direction between tone congruent condition and incongruent condition compared with the Chinese-dominant group ($M= -43, SD=92$), $t(18)=4.078, p=.001$, and the Chinese-English balanced group did ($M= -55, SD=64$), $t(18)=3.995, p=.001$. However, the comparison between the two Chinese groups was not significant, $t(18)= -.479, p=.638$. That is to say, where the two Chinese groups showed the expected tone congruence effect, with enhanced performance for the tone congruent trials, the English group showed the opposite effect, with enhanced performance for the incongruent effect.

Table 3.9. ANOVA results of iRT (overall Study 2B)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	1.078	.308	.038	.17
VC * Language	2	.025	.975	.002	.053
Tone Congruence	1	.312	.581	.011	.084
TC * Language	2	14.81	<.001	.523	.998
VC * TC	1	12.52	.001	.317	.926
VC*TC*Language	2	1.058	.361	.073	.216
Language Group	2	10.479	<.001	.437	.979

Figure 3.7. Tone-angularity congruent condition versus incongruent condition in Study 2B (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

Discussion

When asking participants to pay attention to tones, the non-Chinese group was significantly different from the two Chinese groups in tone-angularity audiovisual congruence effect, whereas the two Chinese groups were not significantly different. The two Chinese groups responded faster to tone-angularity congruent trials (i.e., T1-curved; T4-pointy) than to the incongruent

ones (i.e., T4-curvy; T1-pointy), which was consistent with what has been observed in Study 1 that less-pitch-changing T1 goes with curvy shapes and more-pitch-changing T4 goes with pointy shapes. The results were also consistent with the fact that pitch change is important in Chinese speakers' lexical tone perception (Jongman, et al., 2006; Singh, et al., 2014). On the other hand, the non-Chinese group responded faster to tone-angularity incongruent trials than to the congruent ones. That is, high T1 goes with pointy shapes and low T4 goes with curvy shapes. These results were consistent with the fact that English speakers focus on pitch height in their lexical tone perception (i.e., high-pointy, low-curvy) (Gandour, 1983). According to which, T1, for English speakers, is a better match for the spiky shape, due to its high pitch. In addition, no significant vowel-related group difference was found when tones were attended, suggesting that attention modulated this kind of cross-modal correspondences. These findings (and the power) validate the method for both vowel effect and tone effect with sample sizes as small as 10.

3.4 Study 2C: IAT of Shapes and Vowels in Tone 1 & 3 – Vowel Attended

In the previous section it was established that when tones were attended (i.e., T1, T4), English-speakers with no Chinese had different mappings between tones and shapes from those of the two Chinese groups. Now we attempted to use the same paradigm to test different subject groups' responses to the T1-T3

contrast. Study 2C and 2D were designed to test the non-Chinese participants' lexical tone - angularity mapping pattern (i.e., T1 (high)-pointy; T3 (low)-curvy) using IAT tasks in the three subject groups (C; C/E; E), with vowels attended in Study 2C and tones attended in Study 2D. The procedures of the present study were identical to those of Study 2A, except that instead of the tone contrast T1-T4 used in Study 2A, the tone contrast used here was Mandarin T1-T3. The stimuli used in the present study (2C) were the curvy shape, the pointy shape, /i1/, /i3/, /u1/, /u3/. The manipulation of mappings between stimuli and response keys here was also the same as that in Study 2A, with two angularity-vowel congruent blocks where /u/ and the curvy shape were mapped to one key (either "A" or "L") and /i/ and the pointy shape to the other key, and two angularity-vowel incongruent blocks where /u/ and the pointy shape were mapped to one key and /i/ and the curvy shape to the other key (see Figure 3.3). The tone-angularity congruence used here was the English speakers' mapping pattern observed in Study 1 (i.e., 'high-pointy, low-curvy' mapping pattern. That is, 'T1-curvy, T3-pointy' is incongruent and 'T1-pointy, T3-curvy' is congruent). Note that the mapping instructions in Study 2C were about the relationships among shapes, vowels and keys, and thus tone-angularity congruence was unattended here. The sample sizes of the three groups were 10 (C), 20 (C/E) and 10 (E) respectively.

Results

The accuracy of the three groups was 95.4% (C), 93.2% (C/E) and 94.2% (E) respectively. Data handling was conducted as in study 2A. Figure 3.8 displays the back-transformed RTs shown separately for the three experimental groups, and the different congruence conditions.

Figure 3.8. Bar charts for different subject groups' back-transformed RTs in Study 2C (IAT of Shapes and Vowels in Tone 1 & 3: Vowel Attended) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

Chinese-Dominant Group

The Chinese-dominant group showed a significant main effect of vowel-angularity congruence ($F(1,9)=33.62, p<.001, \eta_p^2=.79$) (see Table 3.10) with no other effects or interactions. Vowel-angularity congruent trials ($M=634, SD=68$) were processed faster than the incongruent ones ($M=687, SD=71$).

Table 3.10. ANOVA results of iRT (Chinese dominant bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	33.62	$p<.001$	.789	.999
Tone Congruence	1	2.35	.159	.207	.279
Interaction	1	1.14	.314	.112	.16

Chinese-English Balanced Bilinguals

The Chinese-English balanced bilinguals also showed a significant main effect of vowel-angularity congruence ($F(1,19)=6.39, p=.02, \eta_p^2=.25$) with no other significant main effects or interactions. Vowel-angularity congruent trials ($M=676, SD=94$) were significantly faster than the incongruent ones ($M=723, SD=148$).

Table 3.11. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	6.39	.02	.252	.67
Congruence					
Tone	1	.021	.886	.001	.052
Congruence					
Interaction	1	2.73	.115	.126	.348

English-speaking Group (non-Chinese)

The English-speaking group showed a significant main effect of vowel-angularity congruence as well ($F(1,9)=6.33, p=.033, \eta_p^2=.41$) (see Table 3.12) with no other significant effects or interactions.. Vowel-angularity congruent trials ($M=766, SD=152$) were significantly faster than the incongruent ones ($M=822, SD=168$).

Table 3.12. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	6.33	.033	.413	.612
Congruence					
Tone	1	.14	.718	.015	.063
Congruence					
Interaction	1	3.64	.089	.288	.4

A comparison among the three groups

To investigate the group difference, as before we ran a 3-way repeated measures ANOVA, with vowel-angularity congruence (congruent, incongruent) and tone-angularity congruence (congruent, incongruent) as within-subjects factors and language group (C, C/E, E) as a between-subjects factor. To keep the equal sample size across language groups ($n=10/\text{group}$), we selected the first 10 participants for the C/E group. The main effect of vowel-angularity congruence was significant ($F(1,27)=27.78, p<.001, \eta_p^2=.507$). Vowel-angularity congruent trials ($M=687, SD=123$) were significantly faster than incongruent ones ($M=744, SD=156$) regardless of tone-angularity congruence effect and this main effect did not differ between the language groups. The interaction between vowel effect and tone effect was also significant ($F(1,27)=10.2, p=.004, \eta_p^2=.274$), but these interactions did not differ

significantly between the levels of language group, therefore we did not further explore them.

Table 3.13. ANOVA results of iRT (overall Study 2C)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	27.78	<.001	.507	.999
VC * Language	2	.586	.564	.042	.137
Tone Congruence	1	.084	.774	.003	.059
TC * Language	2	.854	.437	.059	.181
VC * TC	1	10.2	.004	.274	.868
VC*TC*Language	2	.901	.418	.063	.189
Language Group	2	3.33	.051	.198	.58

Discussion

The ‘*u*-curvy, *i*-pointy’ vowel-angularity audiovisual correspondence has been replicated again using the Mandarin T1-T3 contrast when vowels were attended.

3.5 Study 2D: IAT of Shapes and Vowels in Tone 1 & 3 – Tone Attended

Similar to Study 2C, the present study still used T1-T3 contrast, but instead of vowels, participants were instructed to pay attention to tones and shapes here. The procedures of this study were identical to those of Study 2B. The manipulation of mappings between stimuli and response keys here were based on the English speakers' Mandarin tone-visual angularity mapping pattern. In the two angularity-tone congruent blocks, T3 and the curvy shape were mapped to one key (either “A” or “L”) and T1 and the pointy shape were mapped to the other key. While in the two angularity-tone incongruent blocks, T3 and the pointy shape were mapped to one key and T1 and the curvy shape to the other key (see Figure 3.9). There were ten participants in the Chinese dominant group (C), twenty in the balanced bilingual group (C/E) and ten in the non-Chinese speaker group (E). Since non-Chinese speakers (E) have no knowledge of Mandarin tones, Tone 1 was described as “High sound” and Tone 3 was described as “Low sound” in the instructions for the E Group. The non-Chinese speaking participants were also instructed to listen to the stimuli as much as they wanted, until they were sure that they could correctly identify the High sounds and the Low sounds.

Figure 3.9. Examples of instructions used in the experimental blocks in Study 2D. Left: An example of the instruction for an angularity-tone congruent block. Right: An example of the instruction for an angularity-tone incongruent block.

Results

The accuracy of the three groups was 97% (C), 94% (C/E) and 90% (E) respectively. Data handling was conducted as in study 2B. Figure 3.10 displays the back-transformed RTs shown separately for the three experimental groups, and the different congruence conditions.

Figure 3.10. Bar charts for different subject groups' back-transformed RTs in Study 2D (IAT of Shapes and Vowels in Tone 1 & 3: Tone Attended) (C: Chinese-dominant group; CE: Chinese-English balanced group; E: English group)

Chinese-Dominant Group

No significant effects or interactions were found for this group (see Table 3.14).

Table 3.14. ANOVA results of iRT (Chinese dominant bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	3.65	.089	.288	.4
Congruence					
Tone	1	.051	.826	.006	.055
Congruence					
Interaction	1	3.98	.077	.307	.43

Chinese-English Balanced Bilinguals

A significant interaction effect was found in C/E group (C/E: $F(1,19)=44.9$, $p<.001$, $\eta_p^2=.7$) (see Table 3.15). A paired-samples t-test was conducted to compare the inverse RTs between tone-angularity congruent condition and incongruent condition when vowel and visual angularity were incongruent. There was a significant difference in inverse RTs between tone-angularity congruent condition ($M=828$, $SD=124$) and incongruent condition ($M=735$, $SD=117$) ($t(19) = 5.49$, $p<.001$). Tone-angularity incongruent trials were significantly faster than congruent ones. Tone-angularity crossmodal correspondence appeared when vowels mismatched with visual angularity. A paired-samples t-test was also performed to compare the inverse RTs between

tone-angularity congruent condition and incongruent condition when vowel and visual angularity were congruent. There was no significant difference in inverse RTs between tone-angularity congruent condition ($M=766$, $SD=131$) and incongruent condition ($M=797$, $SD=133$) ($t(19) = -1.53$, $p=.143$).

Table 3.15. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	.001	.973	<.001	.05
Tone Congruence	1	2.9	.105	.132	.366
Interaction	1	44.9	<.001	.703	>.999

English-speaking Group (non-Chinese)

There was a significant main effect of tone-angularity congruence ($F(1,9)=23.78$, $p=.001$, $\eta_p^2=.73$) (see Table 3.16), suggesting that tone congruent trials ($M=865$, $SD=155$) were significantly faster than tone incongruent ones ($M=967$, $SD=162$). The interaction between vowel congruence and tone congruence was also significant ($F(1,9)=9.39$, $p=.013$, $\eta_p^2=.51$). The results of a paired-samples t-test showed that there was a significant difference in inverse RTs between tone-angularity congruent condition ($M=809$, $SD=136$) and incongruent condition ($M=976$, $SD=180$) when vowel and visual angularity were congruent ($t(9)=4.22$, $p=.002$). Tone-angularity congruent trials were significantly faster

than incongruent ones. Tone-angularity congruence effect seemed to be enhanced when vowel and visual angularity were congruent. A paired-samples t-test was also performed to compare the inverse RTs between the tone-angularity congruent and incongruent conditions when vowel and visual angularity were incongruent. There was no significant difference in inverse RTs between tone-angularity congruent ($M=921$, $SD=174$) and incongruent ($M=957$, $SD=143$) conditions ($t(9)=1.55$, $p=.155$).

Table 3.16. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	2.3	.164	.02	.27
Congruence					
Tone	1	23.78	.001	.73	.99
Congruence					
Interaction	1	9.39	.013	.51	.78

A comparison among the three groups

We also ran a 3-way repeated measures ANOVA, with vowel-angularity congruence (congruent, incongruent) and tone-angularity congruence (congruent, incongruent) as within-subjects factors and language group (C, C/E, E) as a between-subjects factor, to test whether the simple effects differ between the levels of language group. To keep the equal sample size across language groups ($n=10/\text{group}$), we selected the first 10 participants for the C/E

group. Along with the previously observed interaction between vowel congruence and tone congruence, there was also a significant interaction between tone-angularity congruence effect and language group ($F(2,27)=7.72$, $p=.002$, $\eta_p^2=.364$). To unpack this interaction, we first computed the difference between the averaged inverse RTs in tone congruent trials and those in tone incongruent trials (i.e., congruent – incongruent) and then we conducted pair-wise independent-samples T Tests between the language groups. The results suggested that the non-Chinese group ($M= -84$, $SD=64$) had effect in different direction between tone congruent condition and incongruent condition from the Chinese-dominant group ($M=3$, $SD=61$), $t(18)=2.85$, $p=.011$, and the Chinese-English balanced group did ($M=36$, $SD=96$), $t(18)=3.39$, $p=.003$. However, the comparison between the two Chinese groups was not significant, $t(18)=1.15$, $p=.264$. These results were consistent with what has been found in Study 2B. The direction of the effect is particularly clear in Figure 3.11.

Table 3.17. ANOVA results of iRT (overall Study 2D)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	.042	.839	.002	.055
VC * Language	2	.752	.493	.051	.16
Tone Congruence	1	1.066	.311	.038	.169
TC * Language	2	7.72	.002	.364	.924
VC * TC	1	13.79	.001	.338	.947
VC*TC*Language	2	1.588	.223	.105	.306
Language Group	2	2.399	.11	.151	.442

Figure 3.11. Tone-angularity congruent condition versus incongruent condition in Study 2D (C: Chinese-dominant group; CE: Chinese-English balanced group; E: English group)

Discussion

When asking participants to pay attention to tones, similar to what has been found in Study 2B, the non-Chinese group was significantly different from the two Chinese groups in tone-angularity audiovisual congruence effect, whereas the two Chinese groups were not significantly different. The Chinese dominant group did not show much difference in reaction time between the tone-

angularity congruent condition and incongruent condition. The Chinese-English bilingual group responded to tone-angularity incongruent trials (T1-curvy; T3-pointy) faster than congruent trials (T1-pointy; T3-curvy). In other words, the Chinese-speaking participants either showed less sensitivity to the high-low tone contrast (i.e., T1-T3) or made responses based on the Chinese pitch-change-driven crossmodal correspondence pattern between lexical tones and visual shape (i.e., less changing T1-curvy; more changing T3-pointy) rather than the English pitch-height-driven crossmodal correspondence pattern between lexical tones and visual shape (i.e., high T1-pointy; low T3-curvy). On the other hand, the non-Chinese group responded to tone-angularity congruent trials (T1-pointy; T3-curvy) faster than incongruent trials (T1-curvy; T3-pointy), which was consistent with English speakers' this kind of crossmodal correspondence pattern (i.e., high-pointy, low-curvy). In addition, the group difference observed here was consistent with what has been found in Study 2B.

Order effect

Order effect might matter in task-switching experiments. Therefore, we checked whether it might matter if participants do a block after a congruent or an incongruent block. The order effect for the Chinese-English balanced group in Study 2A has been examined. Blocks were categorized into four order types: Congruent block after congruent block, Congruent block after incongruent

block, Incongruent block after congruent block, Incongruent block after incongruent. A 2×2 repeated measures ANOVA with vowel-shape congruence and presentation order as within-subjects factors was conducted. As in the primary analysis, the analysis was based on the inverse RTs. The visualization of the data (see Figure 3.12) and the means are reported using back-transformed RTs. The results show that the only significant effect was the vowel-shape congruence effect, as in the primary analyses, ($F(1, 19) = 46.5, p < .001$). Vowel-shape congruent trials ($M = 651, SD = 91$) were much faster than incongruent trials ($M = 789, SD = 114.5$). No significant effect or interaction of presentation order was detected.

Figure 3.12. Order effect (data from Study 2A, Chinese-English balanced group)

3.6. Re-analysis of Study 2 using the D measure

Using the data treatment of inverse transformation coupled with data trimming, we found IAT effects in Study 2. However, Some IAT researchers argue that the D-measure is a more valid test of underlying association strength than alternative ways of handling RT data. In this section, we performed an additional analysis to check if the findings of the original analysis of Study 2 would also be observed using the alternative D measure (For further

information about D see p. 56). Since group differences in treatment effect were detected in Study 2B and 2D in the original analysis, we reported the reanalysis of these two studies first followed by Study 2A and 2C for completeness.

Study 2B (Sound-Shape mapping, T1-T4 contrast, Tone Attended)

In this section, re-analysis of Study 2B was done using the D measure to examine the group differences in the tone-shape congruence effect observed in the original analysis using the inverse RT measure.

Responses greater than three standard deviations above or below the mean were excluded for each participant. Only the RTs for the correct responses were analyzed. According to the guide for calculating the D score (Nosek et al., 2014), the standard deviation of all the RTs in the tone-shape congruent condition and the incongruent condition was computed. The mean of the RTs in the tone-shape congruent condition and the mean of the RTs in the incongruent condition were calculated respectively. D equals to the difference between the two condition means (i.e., congruent condition – incongruent condition) divided by the standard deviation of all the RTs in the two conditions ($D = (M_1 - M_2) / SD$). The D score was calculated for each participant in the three subject groups.

A one-way between-subjects ANOVA showed a significant effect of subject group on the D scores, $F(2,27) = 7.52, p = .003$. Post hoc analyses

(Tukey) indicated that the English speakers were significantly different from the Chinese-dominant group ($p=.004$, $\eta_p^2=1.498$, power=.942) and the Chinese-English balanced group ($p=.012$, $\eta_p^2=1.509$, power=.945), while the two Chinese groups did not significantly differ ($p=.897$, $\eta_p^2=.196$, power=.111) (see Figure 3.13). Both the Chinese-dominant group ($M= -.203$, $SD=.282$) and the Chinese-English balanced group ($M= -.154$, $SD=.214$) were faster in the tone-shape congruent condition than in the incongruent condition. However, the English speakers ($M=.19$, $SD=.241$) were faster in the tone-shape incongruent condition than in the congruent condition. This group difference in D scores was consistent with the group difference in tone-shape congruence effect observed in the original analysis (see p. 110) using the inverse RT measure.

Figure 3.13. Upper: D Scores for each group in Study 2B. (Diamonds represent the mean value for each group); Lower: Individual mean RTs for Congruent and Incongruent Trials

Study 2D (Sound-Shape mapping, T1-T3 contrast, Tone Attended)

The procedure of D calculation was identical to that in Study 2B. A one-way between subjects ANOVA showed a significant effect of subject group on the D scores, $F(2,27) = 5.32, p = .01$. Post hoc analyses (Tukey) indicated that the English speakers were significantly different from the Chinese-dominant group ($p = .042, \eta_p^2 = 1.323, \text{power} = .885$) and the Chinese-English balanced group ($p = .009, \eta_p^2 = 1.245, \text{power} = .85$), while the two Chinese groups did not significantly differ ($p = .622, \eta_p^2 = .38, \text{power} = .204$) (see Figure 3.14). Both the Chinese-dominant group ($M = .018, SD = .181$) and the Chinese-English balanced group ($M = .131, SD = .38$) were faster in the tone-shape incongruent condition than in the congruent condition. However, the English speakers ($M = -.267, SD = .245$) were faster in the tone-shape congruent condition than in the incongruent condition. This group difference in the tone-shape congruence was also consistent with that observed using the inverse RT measure.

Figure 3.14. Upper: D Scores for each group in Study 2D. (Diamonds represent the mean value for each group); Lower: Individual mean RTs for Congruent and Incongruent Trials

Study 2A (Sound-Shape mapping, T1-T4 contrast, Vowel Attended)

In the present study, the mean of the RTs in the vowel-shape congruent condition and the mean of the RTs in the incongruent condition were calculated respectively. The standard deviation of all the RTs in the vowel-shape congruent condition and the incongruent condition was computed. The *D* score was calculated for each participant accordingly in the three subject groups. A one-way between subjects ANOVA did not show a significant effect of subject group on the *D* scores, $F(2,27) = 2.86, p = .754$. All of the participants were faster in the vowel-shape congruent condition than in the incongruent condition without group difference (C: $M = -.297, SD = .255$; CE: $M = -.233, SD = .234$; E: $M = -.222, SD = .222$) (see Figure 3.15). These results were consistent with the findings observed using the inverse RT measure.

Figure 3.15. Upper: D Scores for each group in Study 2A. (Diamonds represent the mean value for each group); Lower: Individual mean RTs for Congruent and Incongruent Trials

Study 2C (Sound-Shape mapping, T1-T3 contrast, Vowel Attended)

The procedure of D calculation was identical to that in Study 2A. A one-way between subjects ANOVA did not show a significant effect of subject group on the D scores either, $F(2,27) = .601, p = .555$. All of the participants were faster in the vowel-shape congruent condition than in the incongruent condition without group difference (C: $M = -.259, SD = .188$; CE: $M = -.174, SD = .258$; E: $M = -.144, SD = .271$) (see Figure 3.16). These results were consistent with the findings observed using the inverse RT measure.

Figure 3.16. Upper: D Scores for each group in Study 2C. (Diamonds represent the mean value for each group); Lower: Individual mean RTs for Congruent and Incongruent Trials

The results observed using the D measure and the inverse RT were consistent throughout Study 2. Hence, the IAT effects observed in Study 2 were

robust. In a sense, the D measure does the same job as the inverse transformation in terms of reducing the skewness of RT distributions caused by slow RTs by using standard deviation as a divisor. Since both methods give essentially the same results, with no appreciable improvement in power, for the remainder of the thesis, the inverse transformation will be used.

3.7 General Discussion

When vowels were attended, for both of the tone contrasts (i.e., T1-T4; T1-T3), all participants responded to vowel-angularity congruent trials (i.e., *u*-curvy, *i*-pointy) significantly faster than incongruent trials (i.e., *u*-pointy, *i*-curvy) and this effect did not differ between language groups. This result replicated the well-documented ‘*u*-curvy, *i*-pointy’ sound-shape crossmodal correspondence pattern (e.g., Köhler 1929:1947; Ramachandran and Hubbard, 2001; Maurer et al., 2006).

Attention modulates crossmodal correspondence patterns. When tones were attended, for both of the tone contrasts, the non-Chinese group was significantly different from the two Chinese groups in tone-angularity congruence effect, whereas the Chinese groups did not significantly differ from each other.

The Chinese groups consistently responded faster to the ‘less pitch changing-curvy, more pitch changing-pointy’ audiovisual mappings (i.e., ‘T1-curvy, T4-pointy’; ‘T1-curvy, T3-pointy’) than to the reverse mappings (i.e.,

‘T1-pointy, T4-curvy’; ‘T1-pointy, T3-curvy’). This efficiency for the ‘steady-curvy, changing-pointy’ crossmodal correspondence was consistent with their preference observed in Study 1, where steady tones go with curvy shapes and changing tones go with pointy shapes. These results were also consistent with the finding that pitch change is important in Chinese speakers’ lexical tone perception (Jongman, et al., 2006; Singh, et al., 2014).

On the other hand, the non-Chinese group whose English was at a native or near-native level responded faster to the ‘high-pointy, low-curvy’ audiovisual mappings (i.e., ‘T1-pointy, T4-curvy’; ‘T1-pointy, T3-curvy’) than to the reverse mappings (i.e., ‘T1-curvy, T4-pointy’; ‘T1-curvy, T3-pointy’). This ‘high-pointy, low-curvy’ crossmodal correspondence pattern was consistent with their preference observed in Study 1, where low tones go with curvy shapes and high tones go with pointy shapes. These results were consistent with the fact that pitch height is important in English speakers’ lexical tone perception (Gandour, 1983).

Taken together, we found no group difference in the ‘*u*-curvy, *i*-pointy’ vowel-shape mapping. However, when it comes to the tone-shape crossmodal correspondence, Chinese speakers showed a pitch-change-driven response pattern, whereas native and near-native English speakers with no Chinese showed a pitch-height-driven response pattern.

Chapter 4: IAT of Shapes and Sine Waves with Tones

4.1 Introduction

In the previous chapter, using lexical tones we demonstrated that participants with different language backgrounds showed different tone effects on tone-angularity mappings. The Chinese groups replicated the pitch change hypothesis observed from the same subject groups in Study 1 (i.e., T1-curvy, T4-pointy, T3-pointy). On the other hand, the non-Chinese English-speaking group showed a pitch-height-driven response pattern by consistently matching the high-pitched T1 with the pointy shape and the low-pitched T3 with the curvy shape.

At this stage, what remains unclear is the sensory underpinnings of these effects. Do the differences in tone-angularity mappings between the groups remain when the pitch contour is presented without the context of a human voice? In this chapter, using IAT tasks, we attempted to investigate whether tone-angularity mapping patterns remain the same for non-linguistic pitch contours, by resynthesizing the sounds used in Study 2 as sine waves. We resynthesized the audio tokens of Tone 1, 3 and 4 from the previous studies, preserving the pitch contour (Study 3A and 3B), and a more subtle variant, which included both the original pitch contour and intensity profile (Study 3C and 3D). As in the previous chapter, the tone contrasts in use here are T1-T4 (Study 3A, 3C) and T1-T3 (Study 3B, 3D).

In Study 2, different subject groups showed different response patterns for this kind of audiovisual cross-modal perception. Will people with different language backgrounds still show different response patterns to the pitch contours of Mandarin Tones when the audio cues of speech are absent? When tested using non-lexical tones, will they use the same tone perception strategy? Or as when tested using lexical tones in Study 2, will they show different strategies (i.e., Chinese speakers: Pitch Change Hypothesis; non-Chinese English-speaking participants: Pitch Height Hypothesis)?

4.2 Study 3A: IAT of Shapes and Sine Waves in Tone 1 & 4

Participants

Twenty participants were recruited for the balanced-bilingual group (C/E) and ten were recruited for the non-Chinese English-speaking group (E). Due to difficulty in recruiting people, we did not have the Chinese-dominant group in Study 3A.

Stimuli

To investigate cross-modal correspondence between visual angularity and Mandarin tones without spectral information, we created sine waves with the pitch architecture of Mandarin T1 and T4 by measuring the pitch of /u1/ and /i4/, and resynthesizing the measured pitch as sine waves using Praat (Boersma & Weenink, 1995). See Figure 4.1 for an example of this manipulation. This

approach preserved pitch architecture of the original recordings, but removed the spectral information of the human voice. The reason /u1/ and /i4/ were used here is that according to Pitch Change Hypothesis, the less changing Tone1 sounds curvy and the more changing Tone4 sounds pointy, meanwhile, ‘*u*-curvy, *i*-pointy’ is a widely reported vowel-angularity matching pattern, and thus /u1/ was assumed to sound curvier than /i1/ and /i4/ was assumed to sound pointier than /u4/. The selection of these two sounds makes T1-T4 contrast more contrastive. The images of the two shapes (one curvy, one pointy) used in Study 1 and Study 2 were still used here.

Figure 4.1. The spectrogram of the original /u3/ used in the study (left) and the spectrogram of the sine wave of it with the measured pitch (right)

Procedure

As in previous studies, participants signed the consent form and did practice trials first. At the beginning of the experimental stage, they were shown an overall instruction, followed by the pictures (one curvy, one pointy) they would see in the experiment. For the two sounds that the participants would hear (one

high; one falling), they were instructed to listen to them one by one as many times as they wanted until they were familiar with which sound was which. Then they could initiate the experiment. Similar to Study 2, at the beginning of each block, participants received the instruction of the mappings between stimuli and response keys for the upcoming block. See Figure 4.2 for the examples of a tone-angularity congruent block (Figure 4.2, Left) and an incongruent block (Figure 4.2, Right). The ‘T1-curvy, T4-pointy’ tone-angularity was based on the Pitch-Change Hypothesis which was a Chinese perception pattern. That is, the less changing T1 goes with the curvy shape and the more pitch changing T4 goes with the pointy shape. We attempted to use the IAT tasks to test this pattern using sine waves, for participants with different language background. The remaining procedure of the task was identical to that of Study 2.

Figure 4.2. Examples of instructions used in the experimental blocks in Study 3A. Left: An example of the instruction for an angularity-tone congruent block. Right: An example of the instruction for an angularity-tone incongruent block.

Results

The accuracy for C/E group and E group were 91.5% and 92% respectively. Data handling was conducted as in study 2. iRTs for both auditory and visual stimuli were analyzed in the present study. In study 2, auditory stimuli had two dimensions (vowel and tone) while visual stimuli had one dimension (angularity). Visual stimuli could only be analyzed in terms of the block-level congruence, and thus excluded from the analysis. However, since angularity-tone congruence was the unique congruence in question here, iRTs of both visual and auditory correct responses were submitted to a repeated-measure analysis of variance (ANOVA) with the within-subjects factors of auditory-visual congruence (auditory versus visual) and tone-angularity congruence (congruent versus incongruent). Figure 4.3 displays the back-transformed RTs shown separately for the two experimental groups, and the different congruence conditions.

Figure 4.3. Bar charts for different subject groups' back-transformed RTs in Study 3A (IAT of Shapes and Sine Waves in Tone 1 & 4) (CE: Chinese-English balanced group; ENG: English group)

Chinese-English Balanced Bilinguals

The bilingual group showed main effect of stimulus modality ($F(1,19)=164.05, p<.001; \eta_p^2=.896$) and main effect of tone-angularity congruence ($F(1,19)=30.14, p<.001, \eta_p^2=.613$) (see Table 4.1), but no significant interaction between them. Overall, visual stimuli ($M=560, SD=92$)

were processed significantly faster than auditory ones ($M=737$, $SD=106$) and angularity-tone incongruent trials ($M=680$, $SD=112$) were faster than congruent ones ($M=608$, $SD=86$).

Table 4.1. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	164.05	<.001	.896	>.999
Tone	1	30.14	<.001	.613	.999
Congruence					
Interaction	1	.814	.378	.041	.138

English-speaking Group (non-Chinese)

Similarly, the English group showed main effect of stimulus modality ($F(1,9)=57.9$, $p<.001$, $\eta_p^2=.87$) and main effect of tone-angularity congruence ($F(1,9)=10.37$, $p=.01$, $\eta_p^2=.54$) (see Table 4.2), but no significant interaction between them. Overall, visual stimuli ($M=588$, $SD=110$) were processed significantly faster than auditory ones ($M=769$, $SD=123$) and angularity-tone incongruent trials ($M=649$, $SD=105$) were faster than congruent ones ($M=708$, $SD=129$).

Table 4.2. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	57.9	<.001	.865	>.999
Tone	1	10.37	.01	.535	.817
Congruence					
Interaction	1	4.16	.072	.316	.445

A comparison between the two groups

To investigate the difference in the effects between the two subject groups, we ran a 3-way repeated measures ANOVA, with stimulus modality (auditory, visual) and tone-angularity congruence (congruent, incongruent) as within-subjects factors and language group (C/E, E) as a between-subjects factor. To keep the equal sample size across language groups ($n=10/\text{group}$), we selected the first 10 participants for the C/E group. The only significant effects were the main effect of stimulus modality ($F(1,18)=139.48, p<.001, \eta_p^2=.89$) and the main effect of tone-angularity congruence ($F(1,18)=25.42, p<.001, \eta_p^2=.59$) (see Table 4.3). Overall, visual stimuli ($M=569, SD=47$) were processed significantly faster than auditory ones ($M=747, SD=111$) and angularity-tone incongruent trials ($M=619, SD=91$) were faster than congruent ones ($M=687, SD=116$). There was no group difference in either of the effects.

Table 4.3. ANOVA results of iRT (overall Study 3A)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Stimulus Modality	1	139.48	<.001	.89	>.999
Modality * Language	1	.12	.738	.006	.062
Tone Congruence	1	25.42	<.001	.585	.997
TC * Language	1	.05	.825	.003	.055
Modality * TC	1	1.82	.194	.092	.248
Modality*TC*Language	1	.15	.708	.008	.0653
Language Group	1	.21	.652	.012	.072

Discussion

For both groups, visual stimuli were faster than audio stimuli and tone-angularity incongruent trials were faster than congruent one. The modality effect may have emerged because, for auditory trials, participants had to wait until the sound file ended which might have caused their responses to auditory trials to be overall slower than visual trials. Moreover, the recognition of angularity might be easier than the recognition of acoustic features, specifically tonal features.

The finding of more theoretical interest here was that for both C/E group and E group, angularity-tone incongruent trials (T1-pointy; T4-curvy) were faster than congruent ones (T1-curvy; T4-pointy). In Study 2B (linguistic T1-

T4 contrast, when tones were attended), the non-Chinese group was significantly different from the two Chinese groups in tone-angularity congruence effect. The two Chinese groups showed the Chinese-style tone-angularity perception pattern (steady (T1)-curvy; changing (T4)-pointy) when responding to linguistic stimuli (/i1/, /i4/, /u1/, /u4/), whereas the non-Chinese group showed a pitch-height-driven ‘high (T1)-pointy, low (T4)-curvy’ pattern. On the contrary, when pitch remained but the spectral information of the human voice was taken away, pitch change no longer influenced the Chinese-English bilinguals in their mappings between sounds and shapes. Instead, the bilinguals, like the English speakers, perceived pitch through pitch height by more effectively matching the on average high T1 with the pointy shape and the on average low T4 with the curvy shape than the reverse mappings.

4.3 Study 3B: IAT of Shapes and Sine Waves in Tone 1 & 3

In this study, we examined cross-modal correspondence between shape and pitch using sine waves more consistent with the English pattern of high/low contrast (i.e., T1-T3 contrast).

Participants

Twenty participants were recruited in C/E group and ten in E group and due to the recruitment difficulty, no Chinese dominant bilingual (C) group is reported here.

Stimuli

We tested T1-T3 contrast in the present study. Sine waves of /i1/ and /u3/ were generated by measuring the dynamic pitch of the audio using PRAAT (Boersma, 2001), and resynthesizing the pitch track as a sine wave, effectively deleting the spectral information from the original audio. The pitch of /i1/ was higher than that of /u1/ and the pitch of /u3/ was lower than that of /i3/. In terms of pitch height, /i1-u3/ is a more typical contrast than other T1-T3 contrasts, and thus they were selected for the generation of sine waves. Spectrograms for sinewave stimuli are provided at Appendix 4.1. The same images were used in the present study.

Procedure

The procedure of the present study was identical to that of Study 3A, except that the tone-angularity congruence here was different from that in Study 3A. In this study, the high sound (T1) and the pointy shape were mapped onto one key and the low sound (T3) and the curvy shape were mapped onto the other key for the congruent blocks, while the high sound (T1) and the curvy shape were

assigned to one key and the low sound (T3) and the pointy shape to the other key for the incongruent blocks (See Figure 4.4). As in the previous study, the participants had to familiarize themselves with the visual images and sounds before they initiated the experiment.

Figure 4.4. Examples of instructions used in the experimental blocks in Study 3B. Left: An example of the instruction for an angularity-tone congruent block. Right: An example of the instruction for an angularity-tone incongruent block.

Results

The accuracy for the two groups was 94.5% (C/E) and 92.2% (E) respectively. Data handling and analysis were conducted as in Study 3A. Figure 4.5 displays the back-transformed RTs shown separately for the two experimental groups, and the different congruence conditions.

Figure 4.5. Bar charts for different subject groups' back-transformed RTs in Study 3B (IAT of Shapes and Sine Waves in Tone 1 & 3) (CE: Chinese-English balanced group; ENG: English group)

Chinese-English Balanced Bilinguals

C/E group showed a main effect of stimulus modality ($F(1,19)=158.4$, $p<.001$, $\eta_p^2=.89$) and a main effect of tone-angularity congruence ($F(1,19)=34.39$, $p<.001$, $\eta_p^2=.64$) (see Table 4.4), but no interaction between them. Like what has been seen in Study 3a, overall, visual stimuli ($M=620$, $SD=136$) were

processed significantly faster than auditory ones ($M=766$, $SD=190$). However, different from the results of Study 3a, angularity-tone congruent trials here ($M=638$, $SD=155$) were significantly faster than incongruent ones ($M=748$, $SD=171$).

Table 4.4. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	158.4	<.001	.893	>.999
Tone	1	34.39	<.001	.644	>.999
Congruence					
Interaction	1	.08	.78	.004	.058

English-speaking Group (non-Chinese)

Like the C/E group, the E group also showed a main effect of stimulus modality ($F(1,9)=10.88$, $p=.009$, $\eta_p^2=.55$) as well as a main effect of tone-angularity congruence ($F(1,9)=11.53$, $p=.008$, $\eta_p^2=.56$) (see Table 4.5), but no interaction between them. Similarly, overall, visual stimuli ($M=666$, $SD=140$) were processed significantly faster than auditory ones ($M=764$, $SD=125$) and angularity-tone congruent trials ($M=680$, $SD=132$) were significantly faster than incongruent ones ($M=750$, $SD=133$).

Table 4.5. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	10.88	.009	.547	.835
Tone	1	11.53	.008	.562	.855
Congruence					
Interaction	1	2.04	.187	.185	.248

A comparison between the two groups

Like in Study 3A, to investigate the difference in the effects between the two subject groups, we ran a 3-way repeated measures ANOVA, with stimulus modality (auditory, visual) and tone-angularity congruence (congruent, incongruent) as within-subjects factors and language group (C/E, E) as a between-subjects factor. Again, to keep the equal sample size across language groups ($n=10/\text{group}$), we selected the first 10 participants for the C/E group. The only significant effects were the main effect of stimulus modality ($F(1,18)=56.51, p<.001, \eta_p^2=.76$) and the main effect of tone-angularity congruence ($F(1,18)=26.12, p<.001, \eta_p^2=.59$) (see Table 4.6). Overall, visual stimuli ($M=651, SD=157$) were processed significantly faster than auditory ones ($M=789, SD=188$) and angularity-tone congruent trials ($M=675, SD=165$)

were significantly faster than incongruent ones ($M=765$, $SD=179$). There was no group difference in either of the effects.

Table 4.6. ANOVA results of iRT (overall Study 3B)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Stimulus Modality	1	56.51	<.001	.76	>.999
Modality * Language	1	2.74	.115	.132	.347
Tone Congruence	1	26.12	<.001	.592	.998
TC * Language	1	.74	.402	.039	.128
Modality * TC	1	.194	.664	.011	.07
Modality*TC*Language	1	3.32	.085	.156	.407
Language Group	1	.062	.807	.003	.056

Discussion

Again, the visual stimuli were responded to significantly faster than the auditory stimuli. In addition, both of the Chinese-English balanced bilingual group and the non-Chinese group responded significantly faster to angularity-tone congruent trials (T1-pointy; T3-curvy) than to incongruent ones (T1-curvy; T3-pointy). It was not surprising to find that the non-Chinese English-speaking participants responded to the ‘T1-pointy, T3-curvy’ mapping faster than to the ‘T1-curvy, T3-pointy’ mapping, since other people from the same subject group

showed the same response pattern when they were presented with lexical tones in Study 2D. By contrast, in the previous chapter, in Study 2D, when using lexical tones with their full spectral information (i.e., /i1/, /i3/, /u1/, /i3/) and asking the Chinese-English balanced bilinguals to pay attention to tones, they responded to the ‘T1-curvy, T3-pointy’ mapping more effectively than to the ‘T1-pointy, T3-curvy’ mapping and the former is a pitch-change-driven Chinese-style angularity-tone mapping pattern. Balanced bilinguals have therefore changed sides by showing the English pitch-height-driven ‘high-pointy, low-curvy’ (i.e., ‘T1-pointy, T3-curvy’) pattern when they were presented with sine wave stimuli.

4.4 Study 3C: IAT of Shapes and Sine Waves with intensity in Tone 1 & 4

In the present study, we applied the measured intensity of /u1/ and /i4/ to the sine waves of /u1/ and /i4/ used in Study 3A respectively using PRAAT to make the perception of the pitch in the sine waves more like the perception of the pitch in the original /u1/ and /i4/ sounds, because in the natural speech, different parts of a tone are loud or soft, and leaving this out of the synthesized stimuli may have artificially inflated the importance of some parts of the pitch contour. Visual stimuli, procedure and the hypothesized tone-angularity congruence were identical to those in Study 3A. There were 10 participants in each participant group (i.e., C, C/E, E).

Results

The accuracy of the three groups was 95.4% (C), 96.2% (C/E) and 92.1% (E) respectively. Data handling and analysis were conducted as in Study 3A. Figure 4.6 displays the back-transformed RTs shown separately for the three experimental groups, and the different congruence conditions.

Figure 4.6. Bar charts for different subject groups' back-transformed RTs in Study 3C (IAT of Shapes and Sine Waves with intensity in Tone 1 & 4) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

Chinese-Dominant Group

Chinese-dominant bilinguals showed a significant main effect of stimulus modality ($F(1, 9) = 143.03, p < .001, \eta_p^2 = .94$) (see Table 4.7). Overall, visual trials ($M = 619, SD = 98$) were processed significantly faster than auditory trials ($M = 799, SD = 93$). Neither the main effect of tone-angularity congruence nor the interaction between stimulus modality and tone congruence was significant (see Table 4.7). On average, tone-angularity incongruent trials were faster than congruent ones, in particular, for auditory trials (see Figure 4.6).

Table 4.7. ANOVA results of iRT (Chinese dominant bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	82.5	<.001	.902	>.999
Tone	1	2.21	.172	.197	.265
Congruence					
Interaction	1	1.39	.268	.134	.1885

Chinese-English Balanced Bilinguals

Chinese-English Balanced Bilinguals showed a significant main effect of stimulus modality ($F(1,9) = 16.99, p = .003, \eta_p^2 = .65$) (see Table 4.8). Overall, visual trials ($M = 558, SD = 154$) were processed significantly faster than auditory trials ($M = 667, SD = 86$). The C/E group also showed a significant main effect of tone-angularity congruence ($F(1,9) = 9.95, p = .012, \eta_p^2 = .53$). Overall, tone-

angularity incongruent trials (M=580, SD=125) were significantly faster than congruent ones (M=656, SD=114). The C/E group also showed a significant interaction between stimulus modality and tone-angularity congruence ($F(1,9)=11.12, p=.009, \eta_p^2=.55$). A paired-samples t-test was conducted to compare the inverse RTs for the auditory trials by the C/E group between tone-angularity congruent condition and incongruent condition. There was no significant difference in the inverse RTs for tone-angularity congruent (M=683, SD=79) and incongruent conditions (M=650, SD=92) ($t(9)= 1.92, p=.087$). A paired-samples t-test was also performed to compare the inverse RTs for the visual trials between tone-angularity congruent condition and incongruent condition. There was a significant difference in the inverse RTs between tone-angularity congruent condition (M=594, SD=149) and incongruent condition (M=512, SD=158) ($t(9)=3.54, p=.006$). Regardless of whether the results of the t-tests were significant or not, tone-angularity incongruent trials were faster than congruent ones for both auditory and visual trials in the C/E group.

Table 4.8. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	16.99	.003	.654	.955
Tone	1	9.95	.012	.525	.801
Congruence					
Interaction	1	11.12	.009	.553	.842

English-speaking Group (non-Chinese)

The E group showed a significant main effect of stimulus modality ($F(1,9)=60.7, p<.001, \eta_p^2=.87$) (see Table 4.9). Overall, visual trials ($M=680, SD=146$) were processed significantly faster than auditory trials ($M=867, SD=203$). The E group also showed a significant main effect of tone-angularity congruence ($F(1,9)=5.72, p=.04, \eta_p^2=.57$). Tone-angularity incongruent trials ($M=739, SD=166$) were faster than congruent ones ($M=795, SD=182$).

Table 4.9. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	60.7	<.001	.871	>.999
Tone	1	5.72	.04	.389	.569
Congruence					
Interaction	1	.03	.867	.003	.053

A comparison among the three groups

To investigate the difference in the effects among the three subject groups, we ran a 3-way repeated measures ANOVA, with stimulus modality (auditory, visual) and tone-angularity congruence (congruent, incongruent) as within-subjects factors and language group (C, C/E, E) as a between-subjects factor. The main effect of stimulus modality ($F(1,27)=101.98, p<.001, \eta_p^2=.79$) and

the main effect of tone-angularity congruence ($F(1,27)=13.87, p=.001, \eta_p^2=.34$) were significant (see Table 4.10). Overall, visual stimuli ($M=619, SD=100$) were processed significantly faster than auditory ones ($M=778, SD=160$) and angularity-tone incongruent trials ($M=678, SD=153$) were faster than congruent ones ($M=739, SD=147$). There was no group difference in either of the effects. The modality-language-tone 3-way interaction and the main effect of language group were also significant (see Table 4.10). However, neither the main effect of language group nor whether the tone-language interactions differ between the levels of stimulus modality was the main concern of the study, and thus they were not further analyzed.

Table 4.10. ANOVA results of iRT (overall Study 3C)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Stimulus Modality	1	101.98	<.001	.79	>.999
Modality * Language	2	.337	.717	.024	.098
Tone Congruence	1	13.87	.001	.339	.948
TC * Language	2	1.83	.181	.119	.347
Modality * TC	1	2.74	.11	.092	.358
Modality*TC*Language	2	6.79	.004	.335	.887
Language Group	2	5.79	.008	.3	.829

Discussion

As in the previous studies in this chapter, visual stimuli were processed faster than auditory stimuli in the present study, probably because the recognition of angularity was easier than that of acoustic features, specifically tonal features. For all groups, angularity-tone incongruent trials (i.e., T1-pointy; T4-curvy) were processed faster than congruent ones (i.e., T4-pointy; T1-curvy) and there was no group difference in this tone-angularity congruence effect, suggesting that all participants responded more effectively to the ‘high-pointy, low-curvy’ mapping than the reverse mapping when the stimuli sounded more like speech sounds but the spectral information was absent. These results were consistent with those of Study 3A, suggesting that all participants perceive non-lexical tones in terms of pitch height. Chinese speakers were no longer Pitch-Change Hypothesis supporters as they were in Study 2B, whereas they showed sensitivity to pitch height and mapped the pointy shape with the on average high T1 and the curvy shape with the low T4. This pattern was the same as what has been observed in people who do not speak a tone language, but different from what has been found in Chinese speakers when presented with lexical tones.

4.5 A comparison between Study 2B & Study 3C

The Mandarin Chinese Tone contrast T1-T4 was used in both Study 2B and Study 3C, with lexical tones (i.e., i1, i4, u1, u4) in Study 2B and sine waves in Study 3C. As seen above, people with different language backgrounds had

different tone-angularity mapping patterns in Study 2B (i.e., Chinese speakers: T1-curvy, T4-pointy; non-Chinese English speakers: T1-pointy, T4-curvy), whereas people with different language backgrounds showed the same tone-angularity mapping pattern (i.e., T1-pointy, T4-curvy) in Study 3C. To test if any difference in the tone-angularity mapping patterns between Study 2B and Study 3C, we ran a 3-way repeated measures ANOVA, with tone-angularity congruence (congruent, incongruent) as a within-subjects factor and audio type (lexical tone, sine wave) and language group (C, C/E, E) as between-subjects factors. All the means compared here were those of auditory stimuli.

Figure 4.7. Comparisons between lexical tones and sine waves (T1-T4) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

There was a significant main effect of tone-angularity congruence ($F(1,54)=7.17, p=.01, \eta_p^2=.117$). Overall, tone-angularity incongruent trials ($M=814, SD=167$) were faster than congruent ones ($M=862, SD=199$), modulated by a significant interaction between audio type and tone effect ($F(1,54)=5.33, p=.025, \eta_p^2=.09$). To unpack this interaction, we first computed the difference between the averaged inverse RTs in tone congruent trials and

those in tone incongruent trials (i.e., congruent – incongruent) and then we conducted an independent-samples T Test between the audio types. The results suggested that the tone effect was moderated by interacting with audio types whereby the tone effect was more prominent in sine waves ($M=63$, $SD=90$) than in lexical tones ($M=18$, $SD=160$), $t(58)=2.93$, $p=.005$. There was also a significant interaction between subject group and tone effect ($F(2,54)=8.2$, $p=.001$, $\eta_p^2=.23$). To unpack this interaction, we computed the difference between the averaged inverse RTs in tone congruent trials and those in tone incongruent trials (i.e., congruent – incongruent) and then we conducted pair-wise independent-samples T Tests between the language groups. The non-Chinese English-speaking group ($M=81$, $SD=89$) showed an overall tone incongruence effect while the Chinese-dominant group ($M=9$, $SD=99$), $t(38)=2.39$, $p=.022$ and the Chinese-English bilingual group ($M=-9$, $SD=69$), $t(38)=3.55$, $p=.001$, did not. Although the two Chinese groups showed no group difference in tone effect, $t(38)=.67$, $p=.5$, this can be understood through the 3-way interaction between tone effect, audio type and subject group which was also significant ($F(2,54)=8.63$, $p=.001$, $\eta_p^2=.242$). As seen in Study 2B, when presented with lexical tones (i.e., i1, i4, u1, u4), the non-Chinese group ($M=123$, $SD=87$) had an inverse tone congruence effect when compared to the Chinese-dominant group ($M=-43$, $SD=92$), $t(18)=4.078$, $p=.001$, and the Chinese-English balanced group ($M=-55$, $SD=64$), $t(18)=3.995$, $p=.001$, and the comparison between the two Chinese groups was not significant, $t(18)=-.479$,

$p=.638$. On the other hand, when it comes to the results for sine waves (i.e., T1, T4) in Study 3C, the main effect of tone-angularity congruence ($F(1,27)=13.87$, $p=.001$, $\eta_p^2=.34$) was significant (angularity-tone incongruent trials were faster than congruent ones), but the interaction between tone congruence and subject group was not significant ($F(2,27)=1.83$, $p=.181$, $\eta_p^2=.119$). Thus, the tone congruence effect across subject groups in different speech types were significantly different, and the observed power (.96) suggested the sample size within each subject group and each audio type was sufficiently large. We also ran independent-samples T Tests between audio types within each subject group. The Chinese-dominant group showed a significant tone congruence difference between lexical tones ($M= -43$, $SD=92$) and sine waves ($M=65$, $SD=82$), $t(18)=2.77$, $p=.013$, with effects in opposite directions across the two audio types. The Chinese-English bilingual group also showed a significant tone congruence effect between lexical tones ($M= -55$, $SD=64$) and sine waves ($M=33$, $SD=63$), $t(18)=3.1$, $p=.006$, also with effects in different directions across the two audio types. The non-Chinese English-speaking group also showed a significant tone congruence difference between lexical tones ($M=123$, $SD=87$) and sine waves ($M=34$, $SD=56$), $t(18)= -2.72$, $p=.015$. However, different from the Chinese groups, the non-Chinese group showed the same direction in the tone-shape congruence effect for both the lexical tones and the sine waves (see Figure. 4.7). The Chinese groups, however, changed the directions of the tone-shape congruence effect depending on the audio types.

Table 4.11. ANOVA results of interactions between Tone Congruence, Speech type and Language group (Study 2B vs. 3C)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Tone Congruence	1	7.17	.01	.117	.749
TC * Audio Type	1	5.33	.025	.09	.621
TC * Language	2	8.2	.001	.233	.951
TC * Audio Type * Language	2	8.63	.001	.242	.96
Audio Type	1	9.13	.004	.145	.843
Language	2	13.59	<.001	.335	.997
Audio Type * Language	2	2.77	.071	.093	.524

Discussion

People with different language backgrounds matched lexical tones with visual angularity differently. Chinese speakers matched them in terms of pitch change, matching less-changing tones (e.g., i1, u1) with curvy shapes and more-changing tones (e.g., i4, u4) with pointy shapes. Non-Chinese English-speaking people matched lexical tones with visual angularity in terms of pitch height, that is, low-pitched tones (e.g., i4, u4) with curvy shapes and high-pitched tones (e.g., i1, u1) with pointy shapes. On the other hand, when it comes to non-lexical sine waves, people, regardless of language backgrounds, matched tones with shapes in terms of pitch height (i.e., high-pitched tones (e.g., T1) with pointy shapes, low-pitched tones (e.g., T4) with curvy shapes). Language backgrounds matter in pitch-shape crossmodal correspondence when people are presented with lexical tones but not with non-lexical sine waves.

4.6 Study 3D: IAT of Shapes and Sine Waves with intensity in Tone 1 & 3

The same type of auditory stimuli as that in Study 3C was used in the present study. To see whether the more English-style contrast (i.e., high-low contrast) makes difference in cross-modal correspondence between shapes and pitch, we used sine waves with intensity in Tone1 and Tone 3 in the present study. We applied the intensity extracted from /i1/ and /u3/ to the T1 and T3 sine waves used in Study 3B respectively Using PRAAT. The visual stimuli, procedure and the hypothesized tone-angularity congruence were identical to

those in Study 3B. There were 10 participants in each participant group (i.e., C, C/E, E).

Results

The accuracy of the three groups was 94.8% (C), 92.5% (C/E) and 92.5% (E) respectively. Data handling and analysis were conducted as in Study 3B. Figure 4.8 displays the back-transformed RTs shown separately for the three experimental groups, and the different congruence conditions.

Figure 4.8. Bar charts for different subject groups' back-transformed RTs in Study 3D (IAT of Shapes and Sine Waves with intensity in Tone 1 & 3) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

Chinese-Dominant Group

As in study 3C, Chinese-Dominant participants showed a significant main effect of stimulus modality ($F(1, 9) = 8.24, p = .018, \eta_p^2 = .48$) (see Table 4.12).

Visual trials ($M = 692, SD = 215$) were performed significantly faster than auditory ones ($M = 782, SD = 145$). This group also showed a significant main

effect of tone-angularity congruence ($F(1, 9) = 10.51, p = .01, \eta_p^2 = .54$). Tone-angularity congruent trials (i.e., T1-pointy; T3-curvy) ($M = 695, SD = 173$) were significantly faster than incongruent ones (i.e. T3-pointy; T1-curvy) ($M = 780, SD = 186$) across groups.

Table 4.12. ANOVA results of iRT (Chinese dominant bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	8.24	.018	.478	.725
Tone	1	10.51	.01	.539	.822
Congruence					
Interaction	1	.009	.925	.001	.051

Chinese-English Balanced Bilinguals

Chinese-English Balanced Bilinguals showed a significant main effect of stimulus modality ($F(1,9) = 42.23, p < .001, \eta_p^2 = .82$) (see Table 4.13). Visual trials ($M = 597, SD = 108$) were performed significantly faster than auditory ones ($M = 760, SD = 88$). This group also showed a significant main effect of tone-angularity congruence ($F(1,9) = 14.77, p = .004, \eta_p^2 = .62$). Tone-angularity congruent trials (i.e., T1-pointy; T3-curvy) ($M = 635, SD = 87$) were significantly faster than incongruent ones (i.e. T3-pointy; T1-curvy) ($M = 720, SD = 109$) across groups.

Table 4.13. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	42.23	<.001	.824	>.999
Tone	1	14.77	.004	.621	.926
Congruence					
Interaction	1	.89	.371	.09	.135

English-speaking Group (non-Chinese)

The English-speaking participants showed a significant main effect of stimulus modality ($F(1,9)=35.35$, $p<.001$, $\eta_p^2=.8$) (see Table 4.14). Visual trials ($M=502$, $SD=79$) were performed significantly faster than auditory ones ($M=628$, $SD=118$). This group also showed a significant main effects of tone-angularity congruence ($F(1,9)=7.29$, $p=.024$, $\eta_p^2=.45$). Tone-angularity congruent trials (i.e., T1-pointy; T3-curvy) ($M=527$, $SD=78$) were significantly faster than incongruent ones (i.e. T3-pointy; T1-curvy) ($M=594$, $SD=119$) across groups.

Table 4.14. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	35.35	<.001	.797	.999
Tone	1	7.29	.024	.448	.672
Congruence					
Interaction	1	.079	.785	.009	.057

Discussion

Similar to the results of Study 3B, when spectral information was absent, all participants responded significantly faster to the ‘high-pitched sounds with pointy shapes, low-pitched sounds with curvy shapes’ mapping than the reverse mapping. No significant difference in this kind of crossmodal correspondence was caused by language backgrounds.

A comparison among the three groups

To investigate the difference in the effects among the three subject groups, we ran a 3-way repeated measures ANOVA, with stimulus modality (auditory, visual) and tone-angularity congruence (congruent, incongruent) as within-subjects factors and language group (C, C/E, E) as a between-subjects factor. The main effect of stimulus modality ($F(1,27)=72.52, p<.001, \eta_p^2=.73$) and the main effect of tone-angularity congruence ($F(1,27)=31.55, p<.001, \eta_p^2=.54$) were significant (see Table 4.15). Overall, visual stimuli ($M=598, SD=164$)

were processed significantly faster than auditory ones ($M=723$, $SD=135$). Angularity-tone congruent trials ($M=620$, $SD=136$) were significantly faster than incongruent ones ($M=694$, $SD=162$). There was no group difference in either of the effects. The main effect of language group was also significant (see Table 4.15). However, it was not the main concern of the study, and thus was not further analyzed.

Table 4.15. ANOVA results of iRT (overall Study 3D)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Stimulus Modality	1	72.52	<.001	.729	>.999
Modality * Language	2	2.09	.144	.134	.391
Tone Congruence	1	31.55	<.001	.539	>.999
TC * Language	2	.13	.879	.009	.068
Modality * TC	1	.76	.392	.027	.134
Modality*TC*Language	2	.17	.848	.012	.073
Language Group	2	7.24	.003	.349	.906

Discussion

Again, visual stimuli were processed much faster than auditory stimuli here. In addition, for all groups, angularity-tone congruent trials (i.e., T1-pointy; T3-curved) were processed much faster than incongruent ones (i.e., T3-pointy; T1-curved) and there was no group difference in this tone-angularity congruence

effect, suggesting that all participants, regardless of language background, responded more effectively to the ‘high-pointy, low-curvy’ mapping than to the reverse mapping when presented with sine wave sounds. In Study 2D, using lexical tones (i1, i3, u1, u3), Chinese speakers did not show the ‘high-pointy, low-curvy’ mapping pattern. However, here, Chinese speakers as well as non-Chinese speakers showed sensitivity to pitch height and mapped the pointy shape with the high T1 and the curvy shape with the low T3 when hearing sine wave sounds. These results were consistent with those of Study 3B, again suggesting that all participants perceive non-lexical tones in terms of pitch height.

4.7 A comparison between Study 2D & Study 3D

Both Study 2D and Study 3D used the Mandarin Chinese Tone contrast T1-T3, with lexical tones (i.e., i1, i3, u1, u3) in Study 2D and sine waves in Study 3D. As seen before, in Study 2D, people with different language backgrounds showed different tone-angularity mapping patterns. That is, the Chinese-speaking participants either showed less sensitivity to the T1-T3 contrast or responded faster to the T1-curvy, T3-pointy mapping than to the reverse mapping. These results were consistent to Chinese speakers’ pitch-changing-driven mapping patterns between lexical tones and visual angularity (i.e., more-changing tones with pointy shapes and less-changing tones with curvy shapes). On the other hand, the non-Chinese English-speaking group responded faster to the pitch-height-driven mapping (i.e., T1-pointy, T3-curvy) than to the reverse

mapping. When it comes to the non-lexical sine wave sounds in Study 3D, however, people with different language backgrounds showed the same tone-angularity mapping pattern (i.e., T1-pointy, T3-curvy). To test if any difference in the tone-angularity mapping patterns between Study 2D and Study 3D, we ran a 3-way repeated measures ANOVA, with tone-angularity congruence (congruent, incongruent) as a within-subjects factor and audio type (lexical tone, sine wave) and language group (C, C/E, E) as between-subjects factors. All the means compared here were those of auditory stimuli.

Figure 4.9. Comparisons between lexical tones and sine waves (T1-T3) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

There was a significant main effect of tone-angularity congruence ($F(1,54)=18, p<.001, \eta_p^2=.25$). Overall, tone-angularity congruent trials ($M=738, SD=146$) were significantly faster than incongruent ones ($M=809, SD=155$). There was also a significant interaction between audio type and tone effect ($F(1,54)=7.16, p=.01, \eta_p^2=.117$). To unpack this interaction, we first computed the difference between the averaged inverse RTs in tone congruent

trials and those in tone incongruent trials (i.e., congruent – incongruent) and then we conducted an independent-samples T Test between the audio types. The tone effect was greater in sine waves ($M = -103$, $SD = 116$) than in lexical tones ($M = -47$, $SD = 152$), $t(58) = 3.51$, $p = .01$. The 3-way interaction between tone effect, audio type and subject group was also significant ($F(2,54) = 7.42$, $p = .001$, $\eta_p^2 = .216$). As seen in Study 2D, when presented with lexical tones (i.e., i1, i3, u1, u3), the non-Chinese group ($M = -84$, $SD = 64$) had a greater tone congruence effect than the Chinese-dominant group ($M = 3$, $SD = 61$), $t(18) = 2.85$, $p = .011$, and the Chinese-English balanced group did ($M = 36$, $SD = 96$), $t(18) = 3.39$, $p = .003$. However, the comparison between the two Chinese groups was not significant, $t(18) = 1.15$, $p = .264$. On the other hand, when it comes to the results for sine waves (i.e., T1, T3) in Study 3D, the main effect of tone-angularity congruence ($F(1,27) = 31.55$, $p < .001$, $\eta_p^2 = .54$) was significant (angularity-tone congruent trials were faster than incongruent ones), but the interaction between tone congruence and subject group was not significant ($F(2,27) = .13$, $p = .879$, $\eta_p^2 = .009$). Thus, there were significantly different patterns of tone congruence effect across subject groups in different audio types, and the observed power (.928) suggested the sample size within each subject group and each audio type was sufficiently large. We also ran independent-samples T Tests between audio types within each subject group. The Chinese-dominant group showed a significant tone congruence difference between lexical tones ($M = 3$, $SD = 61$) and sine waves ($M = -76$, $SD = 108$), $t(18) = 2.39$,

$p=.028$, with almost no effect in the speech condition, and the English direction of effect in the sine condition. The Chinese-English bilingual group also showed a significant tone congruence difference between lexical tones ($M=36$, $SD=96$) and sine waves ($M= -92$, $SD=129$), $t(18)=3.39$, $p=.003$, with opposite directions of the effect for the two audio conditions. On the other hand, the non-Chinese English-speaking group did not show a significant tone congruence effect between lexical tones ($M= -84$, $SD=64$) and sine waves ($M= -57$, $SD=86$), $t(18)= -.622$, $p=.542$. Similar to the results of the comparison between Study 2B and 3C, here, the non-Chinese group showed the similar tone congruence effect to either lexical tones or sine waves, whereas the Chinese groups changed the directions of the tone effect depending on the audio types.

Table 4.16. ANOVA results of interactions between Tone Congruence, Speech type and Language group (Study 2D vs. 3D)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Tone Congruence	1	18	<.001	.25	.986
TC * Audio Type	1	7.16	.01	.117	.748
TC * Language	2	1.36	.266	.048	.28
TC * Audio Type * Language	2	7.42	.001	.216	.928
Audio Type	1	3.84	.055	.066	.486
Language	2	1.74	.184	.061	.35
Audio Type * Language	2	1.33	.272	.047	.276

Discussion

The results here were quite similar to those of the comparison between Study 2B and 3C. People with different language backgrounds matched lexical tones with visual angularity differently. Chinese speakers matched them in terms of pitch change, matching less-changing tones (e.g., i1, u1) with curvy shapes and more-changing tones (e.g., i3, u3) with pointy shapes. Non-Chinese English-speaking people matched lexical tones with visual angularity in terms of pitch height, that is, low-pitched tones (e.g., i3, u3) with curvy shapes and

high-pitched tones (e.g., i1, u1) with pointy shapes. On the other hand, when it comes to non-lexical sine waves, people, regardless of language backgrounds, matched tones with shapes in terms of pitch height (i.e., high-pitched tones (e.g., T1) with pointy shapes, low-pitched tones (e.g., T3) with curvy shapes). It has been proven again that language backgrounds matter in pitch-shape crossmodal correspondence only when people are presented with lexical tones but not when presented with non-lexical sine wave sounds.

4.8 General Discussion

Overall, participants responded faster to visual stimuli than to auditory stimuli, probably because the auditory features were more complicated than the visual features. Also, participants have to wait to find out what the sound is (i.e., it takes time to hear the audio), whereas the entire visual stimulus is available at the same time. Using the same tone contrasts (i.e., T1-T4; T1-T3) and IAT tasks in Study 2 (previous chapter) and Study 3 (this chapter), we got different results for the lexical tones (/i1/, /i3/, /i4/, /u1/, /u3/, /u4/) in Study 2 and for the non-lexical tones (T1, T3, T4) in Study 3 whose pitch and intensity information was identical to that in the lexical tones. After directly comparing the results of lexical tones (Study 2B, 2D) and those of non-lexical sine wave sounds (Study 3), we found that in Study 2B and 2D, when tones were attended, both Chinese dominant bilinguals and Chinese-English balanced bilinguals were faster when more pitch-changing sounds (i.e., /i4/, /u4/, /i3/, /u3/) were mapped with the

pointy shape and less pitch-changing sounds (/i1/, /u1/) mapped with the curvy shape than the reverse mappings, whereas the non-Chinese English-speaking group responded faster to the ‘high tones with pointy shapes, low tones with curvy shapes’ mappings than to the reverse mappings. However, when the spectral information was taken away from the auditory stimuli, regardless of whether the tones are the better simulations of the pitch perception of the original (Study 3C, 3D) or not (Study 3A, 3B), the two Chinese-speaking groups (C & C/E) and the English group (E) no longer showed different response patterns. Instead, pitch height seemed equally important in non-lexical tone perception across subject groups and all participants matched the pointy shape with the on average high-pitched tones and the curvy shape with the on average low-pitched tones. Thus language specificity seems to exist only in the perception of lexical tones but not in the perception of non-lexical tones. It seems like that the perception of non-lexical tones is more basic than that of lexical tones and the perception of lexical tones is more variable or context-dependent. The spectral information of human voice seems to drive the effects observed in the lexical-tone-related crossmodal correspondences.

Chapter 5: IAT of Visual Spatial Frequency and Sounds with Mandarin Chinese Tones

5.1 Introduction

In previous chapters we demonstrated that participants with different language backgrounds showed different response patterns (i.e., Chinese participants' results demonstrated Pitch-Change Hypothesis while English participants' results supported Pitch-Height Hypothesis) to lexical tone stimuli and showed the same response pattern to non-lexical pitch stimuli where the pitch contours were derived from natural articulations of speech, resynthesized onto sine waves (i.e., All participants' results lent support to Pitch-Height Hypothesis when non-lexical tone stimuli were used). Therefore, spectral information of the human voice seems to drive the effects observed in lexical-tone-related sound symbolism, while fundamental frequency (i.e., pitch) is more basic. Regarding the relationship between auditory frequency and visual spatial frequency, Evans and Treisman (2010) found that high auditory frequency was matched with high visual spatial frequency. To see whether the rich detailed complexity of the previous photographs makes a difference in cross-modal mappings between sounds and visual stimuli, in this chapter, we investigate cross-modal correspondence between spatial frequency and Mandarin Chinese tones using both lexical tone stimuli and non-lexical tone stimuli for the first time. We also want to see whether it is still true that language specificity exists only in the perception of lexical tones but not in the

perception of non-lexical tones when the visual dimension of spatial frequency was used.

Spatial frequency is defined as the number of cycles per degree of visual angle. Hence, an image with narrow bars in a sinusoidal grating contains more high spatial frequency information than one with wide bars. The present study attempted to investigate cross-modal correspondence between tones and spatial frequency using IAT tasks. We investigated the cross-modal correspondence generated by fully featured lexical tones in Study 4A, 4B using the stimuli from Study 2A, 2B. We also investigated the cross-modal correspondence generated by non-lexical tones in Study 4C using the stimuli from Study 3C. In order to see the influence of spectral information of the human voice on the cross-modal correspondence between tones and spatial frequency, we compared the results of Study 4B with those of Study 4C. We also compared the cross-modal correspondences between tones and spatial frequency with those between tones and angularity by comparing the results here with those of Chapter 3 and Chapter 4.

5.2 Study 4A: IAT of Visual Spatial Frequency and Vowels in Tone 1 & 4 – Vowel Attended

Participants

There were 10 participants in C group, 20 participants in C/E group and 10 participants in E group.

Stimuli

We used the measured modes of spatial frequency of the two visual stimuli used in our previous studies (one pointy, one curvy) to create Gabor patches using Matlab (R2016a) (see Figure 5.1). The two Gabor patches were sinusoidal gratings with Gaussian envelopes. The size was 400×400 pixels and they were both oriented to the right (135°). The frequency of the two stimuli were 0.0071 cycles/pixel (see Figure 5.1, on the left) and 0.0177 cycles/pixel (see Figure 5.1, on the right) respectively. The image with the low spatial frequency (0.0071 cycles/pixel, see Figure 5.2, on the left) had wider bars in the resulting sinusoidal grating than those of the image with the high spatial frequency (0.0177 cycles/pixel, see Figure 5.2, on the right). We used the same auditory stimuli (i.e., /i1/, /i4/, /u1/, u4/) as in Study 2A. Participants sat approximately 60cm away from the monitor. Visual stimuli were 14.11cm wide and 14.11cm high.

Figure 5.1. Spatial frequency spectrum of the two visual stimuli used in our previous studies (Left: 0.0071 cycles/pixel; Right: 0.0177 cycles/pixel)

Figure 5.2. Visual stimuli used in Study 4 (Left: low spatial frequency; Right: high spatial frequency)

Procedure

The apparatus and procedure here were identical to those in Experiment 2A, except that the instruction of the mappings between the stimuli and the response

keys for the experimental blocks were different from those in Experiment 2A. The spatial frequency of the high-spatial-frequency image was driven from the pointy shape and the spatial frequency of the low-spatial-frequency image was from the curvy shape used in our previous studies. In Study 2A, the pairs of /u/ and the curvy shape, /i/ and the pointy shape were considered congruent, while the pairs of /u/ and the pointy shape, /i/ and the curvy shape were considered incongruent. Therefore, in the present study, in the two spatial frequency-vowel congruent blocks, /u/ and the low-spatial-frequency image (derived from the curvy shape) were assigned to an identical response key (either “A” or “L”) and /i/ and the high-spatial-frequency image (derived from the pointy shape) were assigned to the other response key. Conversely, in the two spatial frequency-vowel incongruent blocks, /u/ and the high-spatial-frequency image were mapped onto an identical response key and /i/ and the low-spatial-frequency image were mapped onto the other key (see Figure 5.3). As in Study 2A, vowel was the attended dimension and tone was the hidden dimension. In Study 2A, the tone-angularity congruence was based on Chinese speakers’ Pitch-Change Hypothesis, that is, the curvy shape was matched with less changing T1 and the pointy shape with more changing T4. The hypothesized tone-angularity congruence here was that T1 with the low-spatial-frequency image, T4 with the high-spatial-frequency image as the congruent condition, while T1 with the high-spatial-frequency image, T4 with the low-spatial-frequency image as the incongruent condition.

Figure 5.3. Examples of instructions used in the experimental blocks in Study 4A. Left: An example of the instruction for a spatial frequency-vowel congruent block. Right: An example of the instruction for a spatial frequency-vowel incongruent block.

Results

The accuracy of the three groups was 93.2% (C), 95.6% (C/E) and 93% (E) respectively. Data handling and analysis were conducted as in Study 2A. Figure 5.4 shows the back-transformed RTs shown separately for the three experimental groups, and the different congruence conditions.

Figure 5.4. Bar charts for different subject groups' back-transformed RTs in Study 4A (IAT of Spatial Frequency and Vowels in Tone 1 & 4: Vowel Attended). (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

Chinese-Dominant Group

The Chinese-dominant bilinguals showed a significant main effect of vowel-spatial frequency congruence ($F(1,9)=6.81$, $p=.028$, $\eta_p^2=.43$) (see Table 5.1). Vowel-spatial frequency congruent trials ($M=747$, $SD=81$) were processed significantly faster than the incongruent ones ($M=834$, $SD=134$).

Table 5.1. ANOVA results of iRT (Chinese dominant bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	6.81	.028	.431	.643
Tone Congruence	1	3.45	.096	.277	.382
Interaction	1	.002	.964	<.001	.05

Chinese-English Balanced Bilinguals

Similarly, the balanced bilinguals also showed a significant main effect of vowel-spatial frequency congruence ($F(1,19)=5.23$, $p=.034$, $\eta_p^2=.22$) (see Table 5.2). Vowel-spatial frequency congruent trials ($M=682$, $SD=91$) were processed faster than the incongruent ones ($M=735$, $SD=108$).

Table 5.2. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	5.23	.034	.216	.583
Tone Congruence	1	.033	.857	.002	.053
Interaction	1	.158	.695	.008	.067

English-speaking Group (non-Chinese)

The non-Chinese English-speaking participants showed a significant main effect of vowel-spatial frequency congruence ($F(1,9)=16.18$, $p=.003$, $\eta_p^2=.64$) (see Table 5.3). Vowel-spatial frequency congruent trials ($M=751$, $SD=146$) were processed significantly faster than the incongruent ones ($M=820$, $SD=153$).

Table 5.3. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	16.18	.003	.643	.946
Tone Congruence	1	.675	.433	.07	.114
Interaction	1	.019	.893	.002	.052

A comparison among the three groups

To investigate the difference in the effects across language groups, we ran a 3-way repeated measures ANOVA, with vowel-spatial frequency congruence (congruent, incongruent) and tone- spatial frequency congruence (congruent, incongruent) as within-subjects factors and language group (C, C/E, E) as a between-subjects factor. To keep the equal sample size across language groups ($n=10/\text{group}$), we selected the first 10 participants for the C/E group. The only significant effect was the main effect of vowel-spatial frequency congruence ($F(1,27)=19.74, p<.001, \eta_p^2=.422$) (see Table 5.4). Vowel-spatial frequency congruent trials ($M=732, SD=108$) were significantly faster than incongruent ones ($M=799, SD=134$) regardless of tone-spatial frequency congruence effect and this main effect did not differ between the language groups.

Table 5.4. ANOVA results of iRT (overall Study 4A)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	19.74	<.001	.422	.99
VC * Language	2	.197	.822	.014	.078
Tone Congruence	1	.001	.98	<.001	.05
TC * Language	2	1.94	.163	.126	.367
VC * TC	1	.008	.929	<.001	.051
VC*TC*Language	2	.041	.96	.003	.056
Language Group	2	1.43	.257	.096	.279

Discussion

When vowels were attended, all participants of the three subject groups showed a significant main effect of the congruence between vowel and spatial frequency (i.e., /u/-low; /i/-high) regardless of whether tones of the vowels were cross-modally congruent with the spatial frequency of the shapes or not. In addition, this main effect did not differ significantly across subject groups. For the first time, we found that high spatial frequency went better with /i/ sound than /u/ sound and low spatial frequency went better with /u/ than /i/ using Gabor patches which only differ in spatial frequency. This finding was novel but not surprising: In Study 2A, /i/ was more matched with the pointy shape with sharp edges containing more high spatial frequency information than the curvy shape without many details containing lower spatial frequency information. Similarly, in the present study, /i/ went better with the image with narrow bars in the grating containing more high spatial frequency information than the image with wide bars in the grating containing lower spatial frequency information. For the same reason for /u/ went better with the low spatial frequency image than the high spatial frequency one.

5.3 Study 4B: IAT of Visual Spatial Frequency and Vowels in Tone 1 & 4 – Tone Attended

Following the logic of Study 2B, in this study, people performed an implicit association test (IAT) where they see Gabor patches with high or low spatial

frequency while listening to vowels articulated in Mandarin Chinese Tones 1 & 4. Tone was the attended dimension and vowel was the hidden dimension.

Participants

There were also ten participants in C group, twenty in C/E group and ten in E group.

Procedure

Stimuli and procedure were identical to those in Study 4A, except that in each block, instead of vowels, one Gabor patch and one *tone* were assigned to one response key (either “A” or “L”) and the other tone and the other image were assigned to the other key. In the present study, the hypothesized tone-spatial frequency congruence was identical to that in Study 4A, that is, in the two spatial frequency-tone congruent blocks, T1 and the low spatial frequency image were assigned to an identical response key (either “A” or “L”) and T4 and the high spatial frequency were assigned to the other response key. Conversely in the two spatial frequency -tone incongruent blocks, T1 and the high spatial frequency were mapped onto an identical response key and T4 and the low spatial frequency were mapped onto the other key (see Figure 5.5).

Figure 5.5. Examples of instructions used in the experimental blocks in Study 4B. Left: An example of the instruction for a spatial frequency-tone congruent block. Right: An example of the instruction for a spatial frequency-tone incongruent block.

Results

The accuracy of the three groups was 95.4% (C), 91% (C/E) and 85.3% (E) respectively. Data handling and analysis were conducted as in Study 2B. Figure 5.6 displays the back-transformed RTs shown separately for the three experimental groups, and the different congruence conditions.

Figure 5.6. Bar charts for different subject groups' back-transformed RTs in Study 4B (IAT of Spatial Frequency and Vowels in Tone 1 & 4: Tone Attended). (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

Chinese-Dominant Group

The Chinese-dominant group showed a significant main effect of tone-spatial frequency congruence ($F(1,9)=20.83, p=.001, \eta_p^2=.7$) (see Table 5.5). Tone-spatial-frequency congruent trials ($M=835, SD=156$) were significantly faster than incongruent ones ($M=957, SD=189$).

Table 5.5. ANOVA results of iRT (Chinese dominant bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	.774	.402	.079	.124
Congruence					
Tone	1	20.83	.001	.698	.981
Congruence					
Interaction	1	.367	.56	.039	.084

Chinese-English Balanced Bilinguals

As in the C group, the C/E group also showed a significant main effect of tone-spatial frequency congruence ($F(1,19)=9.75$, $p=.006$, $\eta_p^2=.34$) (see Table 5.6). Tone-spatial-frequency congruent trials ($M=875$, $SD=192$) were significantly faster than incongruent ones ($M=976$, $SD=197$).

Table 5.6. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	.197	.662	.01	.071
Congruence					
Tone	1	9.75	.006	.339	.842
Congruence					
Interaction	1	4.15	.056	.179	.49

English-speaking Group (non-Chinese)

Group E also showed a significant main effect of tone-spatial frequency congruence ($F(1,9)=14.14$, $p=.004$, $\eta_p^2=.61$) (see Table 5.7). However, different from the two Chinese groups, in the non-Chinese English-speaking group, tone-spatial-frequency incongruent trials ($M=976$, $SD=223$) were significantly faster than congruent ones ($M=1089$, $SD=295$).

Table 5.7. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel	1	.014	.91	.002	.051
Congruence					
Tone	1	14.14	.004	.611	.916
Congruence					
Interaction	1	.221	.65	.024	.071

A comparison among the three groups

We also ran a 3-way repeated measures ANOVA, with vowel-spatial frequency congruence (congruent, incongruent) and tone- spatial frequency congruence (congruent, incongruent) as within-subjects factors and language group (C, C/E, E) as a between-subjects factor, to test whether the simple effects differ between the levels of language group. To keep the equal sample size across language groups ($n=10/\text{group}$), we selected the first 10 participants for the C/E group. The results suggested that the only significant group difference existed in the interaction between tone-spatial frequency congruence and language group ($F(2,27)=10.79$, $p<.001$, $\eta_p^2=.444$). To unpack this interaction, we first computed the difference between the averaged inverse RTs in tone congruent trials and those in tone incongruent trials (i.e., congruent – incongruent) and then we conducted pair-wise independent-samples T Tests

between the language groups. The results suggested that the non-Chinese group ($M=113$, $SD=126$) had a tone effect in the opposite condition to the Chinese-dominant group ($M= -121$, $SD=85$), $t(18)=4.83$, $p<.001$, and the Chinese-English balanced group ($M= -109$, $SD=167$), $t(18)=3.35$, $p=.004$. However, the comparison between the two Chinese groups was not significant, $t(18)= -.2$, $p=.844$. And this difference in direction can be seen clearly in figure 5.7.

Table 5.8. ANOVA results of iRT (overall Study 4B)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Vowel Congruence	1	1.41	.245	.05	.209
VC * Language	2	.254	.777	.019	.086
Tone Congruence	1	2.67	.114	.09	.35
TC * Language	2	10.79	<.001	.444	.982
VC * TC	1	.717	.405	.026	.129
VC*TC*Language	2	.095	.91	.007	.063
Language Group	2	1.04	.367	.072	.213

Figure 5.7. Tone-angularity congruent condition versus incongruent condition in Study 4B (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

Discussion

Through a direct comparison between subject groups, we observed that when linguistic stimuli (i.e., /i1/, /i4/, /u1/, u4/) were used and tones were attended, both Chinese-speaking groups performed significantly more effectively to ‘more changing T4-high spatial frequency, less changing T1-low spatial frequency’ mappings than ‘more changing T4-low spatial frequency,

less changing T1-high spatial frequency' mappings. Overall, this finding is inconsistent with Evans and Treisman (2010)'s finding that high spatial frequency was paired with high pitch (T1 was higher than T4 here) and low spatial frequency with low pitch (T4 was lower than T1 here). Our results demonstrated that Chinese speakers' sensitivity to pitch change mainly worked for the cross-modal correspondence between Mandarin tones and spatial frequency. On the other hand, the non-Chinese English-speaking group responded more effectively to the 'high-high, low-low' mapping pattern (i.e., on average high T1 with high-spatial-frequency images and on average low T4 with low-spatial-frequency images) than to the reverse mapping, which is consistent with Evans and Treisman (2010)'s finding. These results again suggested that pitch height is important in English speakers' cross-modal correspondence between lexical tones and spatial frequency. Moreover, both study 2B and this study showed the similar group difference between the Chinese groups and the English-speaking non-Chinese group in the audiovisual mapping patterns. That is, pitch change is important in Chinese speakers' cross-modal correspondences between visual dimensions (i.e., angularity, spatial frequency) and lexical tones (i.e., Mandarin Chinese tones) and pitch height is important in English speakers' cross-modal correspondences.

5.4 Study 4C: IAT of Visual Spatial Frequency and Sine Waves with intensity in Tone 1 & 4

In the previous studies, we compared visual spatial frequency and lexical tones. Here, we investigated the cross-modal correspondence between visual spatial frequency (derived from photographs of complex 3-D objects, but missing the spectral context of the rich visual shape and texture) and non-linguistic pitch contours of Tone 1 & 4 (derived from natural speech, but missing the spectral context of human voice) to see whether the spectral context of human voice makes difference in cross-modal correspondence between visual spatial frequency and tones.

Participants

There were ten participants in C group, twenty participants in C/E group and ten participants in E group.

Procedure

The auditory stimuli were those used in Study 3C, which were obtained by synthesizing the measured pitch and intensity of /u1/ and /i4/ as sine waves. The procedure was the same as that in Study 3C. Visual stimuli and the hypothesized tone-spatial frequency congruence were the same as those in Study 4B. According to the conclusion of Study 4B (i.e., the greater the pitch change within a Mandarin tone, the higher the spatial frequency of its association), in the two spatial frequency-tone congruent blocks, less changing

T1 and the low spatial frequency image were assigned to an identical response key (either “A” or “L”) and more changing T4 and the high spatial frequency were assigned to the other response key. While in the two spatial frequency-tone incongruent blocks, T1 and the high spatial frequency were assigned to an identical response key, and T4 and the low spatial frequency were assigned to the other key (see Figure 5.8).

Figure 5.8. Examples of instructions used in the experimental blocks in Study 4C. Left: An example of the instruction for a spatial frequency-tone congruent block. Right: An example of the instruction for a spatial frequency-tone incongruent block.

Results

The accuracy of the three groups was 93.5% (C), 93% (C/E) and 95.2% (E) respectively. Data handling and analysis were conducted as in Study 3C. Figure 5.9 displays the back-transformed RTs shown separately for the three experimental groups, and the different congruence conditions.

Figure 5.9. Bar charts for different subject groups' back-transformed RTs in Study 4C (IAT of Spatial Frequency and Sine Waves with intensity in Tone 1 & 4). (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

Chinese-Dominant Group

The Chinese-Dominant group showed a significant main effect of tone-spatial frequency congruence ($F(1,9)=5.42$, $p=.045$, $\eta_p^2=.37$) (see Table 5.9).

Tone-spatial frequency incongruent trials (i.e., T1-high; T4-low) (M=706, SD=165) were processed faster than congruent ones (i.e., T1-low; T4-high) (M=753, SD=134). A main effect of stimulus modality was also significant for this group ($F(1,9)=20.33$, $p=.001$, $\eta_p^2=.7$). Visual stimuli (M=652, SD=130) were processed much faster than auditory stimuli (M=797, SD=169).

Table 5.9. ANOVA results of iRT (Chinese dominant bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	20.33	.001	.693	.979
Tone	1	5.42	.045	.376	.546
Congruence					
Interaction	1	.078	.787	.009	.057

Chinese-English Balanced Bilinguals

The Chinese-English Balanced Bilingual group showed a significant main effect of tone-spatial frequency congruence effect ($F(1,19)=7.53$, $p=.013$, $\eta_p^2=.28$) (see Table 5.10). Tone-spatial frequency incongruent trials (i.e., T1-high; T4-low) (M=663, SD=86) were processed faster than congruent ones (i.e., T1-low; T4-high) (M=738, SD=137). A main effect of stimulus modality was also significant for this group ($F(1,19)=63.03$, $p<.001$, $\eta_p^2=.79$). Visual stimuli

(M=641, SD=118) were processed much faster than auditory stimuli (M=771, SD=105).

Table 5.10. ANOVA results of iRT (Chinese-English balanced bilinguals)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	63.03	<.001	.786	>.999
Tone	1	7.53	.013	.284	.74
Congruence					
Interaction	1	.41	.531	.021	.093

English-speaking Group (non-Chinese)

The English-speaking group showed a significant main effect of tone-spatial frequency congruence ($F(1,9)=5.56, p=.043, \eta_p^2=.38$) (see Table 5.11). Tone-spatial frequency incongruent trials (i.e., T1-high; T4-low) ($M=711, SD=152$) were processed much faster than congruent ones (i.e., T1-low; T4-high) ($M=800, SD=85$). A main effect of stimulus modality was also significant for this group ($F(1,9)=33.64, p<.001, \eta_p^2=.79$). Visual stimuli ($M=667, SD=129$) were processed much faster than auditory stimuli ($M=844, SD=108$).

Table 5.11. ANOVA results of iRT (non-Chinese English-speaking participants)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Modality	1	33.64	<.001	.789	.999
Tone	1	5.56	.043	.382	.557
Congruence					
Interaction	1	1.63	.234	.153	.208

A comparison among the three groups

To investigate the difference in the effects between the two subject groups, we ran a 3-way repeated measures ANOVA, with stimulus modality (auditory, visual) and tone-spatial frequency congruence (congruent, incongruent) as within-subjects factors and language group (C/E, E) as a between-subjects factor. To keep equal sample sizes across language groups ($n=10/\text{group}$), we selected the first 10 participants for the C/E group. The only significant effects were the main effect of stimulus modality ($F(1,27)=83.1, p<.001, \eta_p^2=.755$) and the main effect of tone-spatial frequency congruence ($F(1,27)=13.32, p=.001, \eta_p^2=.33$) (see Table 5.12). Overall, visual stimuli ($M=655, SD=118$) were processed significantly faster than auditory ones ($M=804, SD=113$) and tone-spatial frequency incongruent trials ($M=674, SD=114$) were faster than congruent ones ($M=763, SD=136$). There was no group difference in either of the effects.

Table 5.12. ANOVA results of iRT (overall Study 4C)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Stimulus Modality	1	83.1	<.001	.755	>.999
Modality * Language	2	.227	.798	.017	.082
Tone Congruence	1	13.32	.001	.33	.94
TC * Language	2	.113	.894	.008	.066
Modality * TC	1	2.79	.107	.094	.363
Modality*TC*Language	2	.542	.588	.039	.13
Language Group	2	.609	.551	.043	.141

Discussion

In all groups, visual stimuli were processed significantly faster than audio stimuli without significant group difference in this effect. This sensory modality effect has also been found in our other sine-wave studies. It is probably because the recognition of spatial frequency is easier than that of acoustic features, specifically tonal features. In addition, in all groups, tone-spatial frequency incongruent trials (T1-high; T4-low) were faster than congruent ones (T1-low; T4-high) and there was no significant group difference in this effect of congruence between tone and spatial frequency. Remember, in Study 4B (linguistic T1-T4 contrast, when tones were attended), the non-Chinese group

was significantly different from the two Chinese groups in the tone-spatial frequency congruence effect. The two Chinese groups showed the Chinese-style cross-modal correspondence pattern between tone and spatial frequency (i.e., steady (T1)-low; changing (T4)-high) when responding to linguistic stimuli (/i1/, /i4/, /u1/, /u4/), whereas the non-Chinese group showed a pitch-height-driven ‘high (T1)-high, low (T4)-low’ pattern. On the contrary, when pitch remained but the spectral information of the human voice was taken away, pitch change no longer influenced the Chinese speakers’ mappings between sounds and shapes. Instead, the Chinese speakers, like the English speakers, perceived pitch through pitch height by more effectively matching the on average high T1 with the shape with high spatial frequency and the on average low T4 with the shape with low spatial frequency than the reverse mappings. This finding is consistent with the previous findings of Evans and Treisman (2010). These results are also quite similar to what have been seen in our previous studies with visual stimuli differing in visual angularity.

5.5 A comparison between Study 4B & Study 4C

The Mandarin Chinese Tone contrast T1-T4 was used in both Study 4B and Study 4C, with lexical tones (i.e., i1, i4, u1, u4) in Study 4B and sine waves in Study 4C. In Study 4B, when presented with lexical tones, people with different language backgrounds had different mapping patterns between tone and spatial frequency (i.e., Chinese speakers: T1-low, T4-high; non-Chinese English

speakers: T1-high, T4-low). However, in Study 4C, when spectral information was taken away from the audio, all participants, regardless of language background showed the same pitch-height-driven response pattern, by responding to ‘high pitch (T1)-high spatial frequency, low pitch (T4)-low spatial frequency’ mapping faster than ‘high pitch (T1)-low spatial frequency, low pitch (T4)-high spatial frequency’ mapping. To test if any statistically significant difference in the pitch-spatial frequency mapping patterns between Study 4B and Study 4C, we ran a 3-way repeated measures ANOVA, with tone-spatial frequency congruence (congruent, incongruent) as a within-subjects factor and audio type (lexical tone, sine wave) and language group (C, C/E, E) as between-subjects factors. All the means compared here were those of auditory stimuli.

Figure 5.10. Comparisons between lexical tones and sine waves (T1-T4) (C: Chinese-dominant group; CE: Chinese-English balanced group; ENG: English group)

There was also a significant interaction between audio type and tone effect ($F(1,54)=18.91, p<.001, \eta_p^2=.259$). To unpack this interaction, we first computed the difference between the averaged inverse RTs in tone congruent trials and those in tone incongruent trials (i.e., congruent – incongruent) and then we conducted an independent-samples T Test between the audio types. The results suggested that the tone effect was moderated by interacting with

audio types whereby the tone effect was more prominent in sine waves ($M=63$, $SD=110$) than in lexical tones ($M= -41$, $SD=160$), $t(58)=2.94$, $p=.005$. There was also a significant interaction between subject group and tone effect ($F(2,54)=5.85$, $p=.005$, $\eta_p^2=.178$). To unpack this interaction, we computed the difference between the averaged inverse RTs in tone congruent trials and those in tone incongruent trials (i.e., congruent – incongruent) and then we conducted pair-wise independent-samples T Tests between the language groups. The non-Chinese English-speaking group ($M=99$, $SD=127$) showed more tone effect than the Chinese-dominant group ($M= -43$, $SD=99$), $t(38)=2.86$, $p=.007$ and the Chinese-English bilingual group ($M= -24$, $SD=156$), $t(38)=2.71$, $p=.01$. The two Chinese groups showed no group difference in tone effect, $t(38)=.45$, $p=.654$. The 3-way interaction between tone effect, audio type and subject group was also significant ($F(2,54)=4.1$, $p=.022$, $\eta_p^2=.132$). As seen in Study 4B, when presented with lexical tones (i.e., i1, i4, u1, u4), the non-Chinese group ($M=113$, $SD=126$) had difference in the opposite direction to the Chinese-dominant group ($M= -121$, $SD=85$), $t(18)=4.83$, $p<.001$, and the Chinese-English balanced group did ($M= -109$, $SD=167$), $t(18)=3.35$, $p=.004$. However, the comparison between the two Chinese groups was not significant, $t(18)= -.2$, $p=.844$. On the other hand, when it comes to the results for sine waves (i.e., T1, T4) in Study 4C, the main effect of tone-spatial frequency congruence ($F(1,27)=13.32$, $p=.001$, $\eta_p^2=.33$) was significant (tone-spatial frequency incongruent trials were faster than congruent ones), but the

interaction between tone congruence and subject group was not significant ($F(2,27)=.113, p=.894, \eta_p^2=.008$). Thus, the tone congruence effect across subject groups in different audio types were significantly different. We also ran independent-samples T Tests between audio types within each subject group. The Chinese-dominant group showed a significant tone congruence effect between lexical tones ($M= -121, SD=88$) and sine waves ($M=39, SD=84$), $t(18)=4.16, p=.001$. The Chinese-English bilingual group also showed a significant tone congruence effect between lexical tones ($M= -109, SD=166$) and sine waves ($M= 63, SD=91$), $t(18)=2.84, p=.011$. On the other hand, the non-Chinese English-speaking group did not show a significant tone congruence effect between lexical tones ($M=113, SD=110$) and sine waves ($M= 93, SD=150$), $t(18)=.33, p=.745$. The non-Chinese group showed the similar tone congruence effect to either lexical tones or sine waves. Whereas the Chinese groups changed the directions of the tone effect depending on the audio types.

Table 5.13. ANOVA results of interactions between Tone Congruence, Speech type and Language group (Study 4B vs. 4C)

Source	df	F	<i>p</i>	Partial η^2	Observed Power
Tone Congruence	1	.979	.327	.018	.163
TC * Audio Type	1	18.91	<.001	.259	.99
TC * Language	2	5.85	.005	.178	.855
TC * Audio Type * Language	2	4.1	.022	.132	.703
Audio Type	1	12.6	.001	.189	.937
Language	2	2.07	.137	.071	.407
Audio Type * Language	2	.13	.878	.005	.069

Discussion

When presented with non-lexical sine wave sounds, all people, regardless of language backgrounds, showed the same response pattern of mappings between non-lexical tones and spatial frequency as that documented in Evans and Treisman's study (2010), by pairing high spatial frequency with high pitch and low spatial frequency with low pitch. This pitch-height-driven response pattern was different from the pitch-change-driven one (i.e., more changing T4 with

high spatial frequency, less changing T1 with low spatial frequency) observed in the two Chinese groups (C; C/E) in Study 4B where lexical tones were used. The Chinese speakers, as tonal language speakers, changed their mappings patterns between pitch and spatial frequency depending on audio types, whereas the non-Chinese English speakers kept the same mapping pattern across audio types. Language backgrounds seem to matter in cross-modal correspondences between lexical tones and spatial frequency only but not in those between non-lexical tones and spatial frequency. Thus, it seems like that lexical tone processing is specific to the context of the language while non-lexical tone processing is more basic and universal.

5.6 General Discussion

In Study 4A, we demonstrated the ‘*u*-low, *i*-high’ vowel-spatial frequency correspondence for the first time using Gabor patches (extremely pure representations of the modal spatial frequency of the visual stimuli used in our previous studies). This finding was consistent with the ‘*u*-curvy, *i*-pointy’ vowel-angularity mapping pattern found in 2A, as the curvy shape contained low spatial frequency information and the pointy shape contained high spatial frequency information. Moreover, as seen in both tone-angularity correspondence in Study 2B and tone-spatial-frequency correspondence in Study 4B, when lexical tones were used and tones were attended, both Chinese groups showed pitch-change-driven response patterns. That is, in Study 2B, the

Chinese groups (i.e., C & CE) showed ‘more-pitch-changing tones paired with pointy shapes, less-pitch-changing tones paired with curvy shapes’ pattern and in Study 4B, both Chinese groups paired more-pitch-changing tones with high spatial frequency images and less-pitch-changing tones with low spatial frequency images. Such results were caused by the fact that pitch change is important in Chinese speakers’ lexical tone perception (see Jongman, et al., 2006; Singh, et al., 2014 for more discussion on this point). However, the non-Chinese English speakers showed the pitched-height-driven mapping pattern in both 2B and 4B (i.e., 2B: high-pointy, low-curvy; 4B: high-high, low-low). In addition, the difference in the crossmodal mapping patterns across subject groups between audio types (i.e., lexical tones vs. sine wave sounds) was significant. On the other hand, when non-lexical tones were used in Study 3C, 3D and 4C, all participants, regardless of language backgrounds, showed the ‘high-pointy, low-curvy’ mapping for tone-angularity correspondence in Study 3C and 3D and ‘high-high, low-low’ mapping for tone-spatial-frequency correspondence in Study 4C which was consistent with the English speakers’ mapping between pure tones and spatial frequency (Evans and Treisman, 2010), suggesting that all people seemed to direct their attention to pitch height of non-lexical tones. Such results were also consistent with English speakers’ pitch-height-focused pitch perception pattern (Gandour, 1983). The results also demonstrated that language specificity in both pitch perception and tone-related audiovisual cross-modal correspondence disappeared when non-lexical tones were used.

Chapter 6: Conclusion

The ‘*u*-curvy, *i*-pointy’ vowel-angularity cross-modal correspondence has been observed across cultures in adults and developmentally in toddlers and young infants (Köhler, 1929, 1947; Davis, 1961; Holland and Wertheimer’s studies, 1964; Tarte and Barritt, 1971; Tarte, 1974; Ramachandran and Hubbard, 2001; Maurer et al., 2006; Nielsen and Rendall, 2011; 2013; Bremner et al., 2013; Ozturk et al., 2013; D’Onofrio, 2014). Additionally, there has been some evidence regarding cross-modal correspondence between pure tones and angularity consistently reporting the ‘high-pointy, low-curvy’ mapping pattern (Marks, 1974; 1987; Walker et al., 2010). What about the relationship between visual angularity and tones in speech sounds, for example, Mandarin tones? Until this body of research was undertaken, there was no research investigating the relationship between lexical tones and visual dimensions. In the present investigations, in Chapter 2, we examined cross-modal mappings between Mandarin tones and shapes using 2AFC tasks. Chapter 3 and Chapter 4 investigated mappings between Mandarin tones and shapes and mappings between shapes and pitch extracted from human voice respectively using IAT tasks. Chapter 5 explored cross-modal correspondence between spatial frequency and lexical/non-lexical tones using IAT tasks. In all of these studies we found that Chinese-speaking participants showed mappings for pitch change, while non-Chinese-speaking participants showed mappings for pitch height, when the stimuli were linguistic, in particular, when tones were

attended. On the other hand, all participants regardless of language backgrounds showed pitch-height-oriented mappings when non-linguistic stimuli were used.

6.1 Language specificity and language acquisition

Kuhl's (Kuhl, 2004; Kuhl et al., 2006) experiments have shown that infants between 6-8 months can hear a universal set of sound contrasts that may come from any language. And two months later, between 8-10 months, most infants become committed to the sounds they hear most often, coupled with loss of sensitivity to those sounds which are heard infrequently due to experience-dependent plasticity. Kuhl et al. (2006) argued that infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Studies on infants' tone perception have also lent support to this perceptual reorganization hypothesis by finding that 4-6-month-old English infants showed comparable sensitivity to Thai tones with that of Chinese infants at the same age. However, English infants' sensitivity to Thai tones declined from 6 to 9 months while Chinese infants' sensitivity to Thai tones remained the same over the same time period (Mattock & Burnham, 2006; Mattock et al., 2008). Additionally, perception of lexical tones has been proven to be different in adults with different language backgrounds. Fundamental frequency variation (i.e., pitch change) is important in Mandarin adult listeners' lexical tone perception (Jongman, et al., 2006; Singh, et al., 2014), while English adult listeners pay the most attention to pitch height (Gandour, 1983). In this thesis,

we proposed that perceptual changes caused by language learning might cause other changes in the perceptual system, for example, changing the nature of cross-modal perception.

6.2 Cross-modal correspondence between vowels, visual angularity and Mandarin tones

2AFC tasks (Study 1A & 1B) using lexical tones

In these two studies, participants were asked to choose one from two options (images or sounds) for one stimulus (auditory or visual). Overall, in terms of vowel-visual angularity cross-modal correspondence, the ‘*u*-curvy, *i*-pointy’ mapping pattern was found in all three participant groups (C; C/E; E) replicating the results in the literature. In terms of visual angularity-Mandarin tone cross-modal correspondence, participants with different language background showed distinct response patterns. English participants (E) matched high-pitched sounds with pointy shapes and low-pitched sounds with curvy shapes, which was consistent with the fact that English listeners pay the most attention to pitch height when perceiving lexical tones. Chinese dominant bilinguals (C), on the other hand, matched more-pitch-changing sounds with pointy shapes and less-changing sounds with curvy shapes, which was also consistent with Chinese listeners’ focus of lexical tone perception — fundamental frequency variation (i.e., pitch change). Pitch-Change Hypothesis, that is, the greater the pitch change within a Mandarin tone, the higher the spatial frequency of its association, has been demonstrated by the Chinese

dominant bilinguals. While Chinese-English balanced bilinguals (C/E) showed a Chinese pitch-change-driven response pattern in Study 1A, but showed an English-like pitch-height-driven mapping pattern in Study 1B when presented with a pair of tones where vowels were controlled. The balanced bilinguals seemed to swing between the two strategies depending on task types. This might be mainly due to their fairly balanced proficiency in both Chinese and English. Moreover, the vowel effect was stronger than tone effect across participant groups, which was consistent with the previous literature (Tong et al., 2008). To sum up, this study is the first to show that language background shapes the cross-modal perception of tones. One possible explanation for this pattern of effects is that language background tunes multisensory perception between visual angularity and lexical tones. This finding could help to explain why in Lapolla (1994) English speakers were better at mappings between Mandarin tones and word meanings after tones were switched.

IAT tasks using lexical tones (Study 2A, 2B, 2C, 2D)

2AFC tasks can lead to limited number of choices and ‘strategic’ responses when participants are asked to think about the relationship between the shape and the sound. Implicit Association Test (IAT) tasks can resolve the above-mentioned methodology limitations by using indirect measure of cross-modal correspondence (Parise & Spence, 2012). When vowels were attended, the vowel-angularity congruence effect was significant across all the participant groups for both T1-T4 contrast and T1-T3 contrast. The well documented ‘*u*-

curvy, *i*-pointy' crossmodal correspondence has been replicated again using IAT tasks (e.g., Köhler 1929:1947; Ramachandran and Hubbard, 2001; Maurer et al., 2006; Nielsen and Rendall, 2011; 2013; D'Onofrio, 2014). Attention modulates crossmodal correspondence patterns. When tones were attended for the T1-T4 contrast in Study B, the results of the group comparison showed that the non-Chinese group was significantly different from the two Chinese groups in tone-angularity audiovisual congruence effect, whereas the two Chinese groups were not significantly different. The two Chinese groups responded faster to tone-angularity congruent trials (i.e., T1-curvy; T4-pointy) than to the incongruent ones (i.e., T4-curvy; T1-pointy), which was consistent with what has been observed in Study 1 that less-pitch-changing T1 goes with curvy shapes and more-pitch-changing T4 goes with pointy shapes. The results were also consistent with the fact that pitch change is important in Chinese speakers' lexical tone perception (Jongman, et al., 2006; Singh, et al., 2014). This finding here was also consistent with some recent findings on pitch processing at the level of the brainstem and cortex (Bidelman & Chung, 2015; Chandrasekaran, Gandour, & Krishnan, 2007; Chandrasekaran, Sampath, & Wong, 2010; Gandour & Krishnan, 2014, 2016; Krishnan & Gandour, 2009; Krishnan, Gandour, & Suresh, 2015, 2016; Krishnan et al., 2009; Krishnan, Xu, Gandour, & Cariani, 2005; Prom-on, Xu, & Thipakorn, 2009; Swaminathan, Krishnan, & Gandour, 2008; Xu, 2001). On the other hand, the non-Chinese English-speaking group responded faster to tone-angularity incongruent trials than to the congruent ones. That is, high T1 goes with pointy shapes and low T4 goes

with curvy shapes. These results are consistent with the fact that English speakers focus on pitch height in their lexical tone perception (i.e., high-pointy, low-curvy) (Gandour, 1983). In addition, no significant vowel-related group difference was found when tones were attended, suggesting that attention modulated this kind of cross-modal correspondences. When T1 and T3 were attended in Study 2D, similar to what has been found in Study 2B, the non-Chinese group was significantly different from the two Chinese groups in tone-angularity audiovisual congruence effect, whereas the two Chinese groups were not significantly different. The Chinese dominant group did not show much difference in reaction time between the tone-angularity congruent condition and incongruent condition. The Chinese-English bilingual group responded to tone-angularity incongruent trials (T1-curvy; T3-pointy) faster than congruent trials (T1-pointy; T3-curvy). In other words, the Chinese-speaking participants either showed less sensitivity to the high-low tone contrast (i.e., T1-T3) or made responses based on the Chinese pitch-change-driven crossmodal correspondence pattern between lexical tones and visual shape (i.e., less changing T1-curvy; more changing T3-pointy) rather than the English pitch-height-driven crossmodal correspondence pattern between lexical tones and visual shape (i.e., high T1-pointy; low T3-curvy). On the other hand, the non-Chinese group responded to tone-angularity congruent trials (T1-pointy; T3-curvy) faster than incongruent trials (T1-curvy; T3-pointy), which was consistent with English speakers' this kind of crossmodal correspondence pattern (i.e., high-pointy, low-curvy). In conclusion, Chinese speakers

consistently showed a pitch-change-driven response pattern in both Study 1 and here, whereas native and near-native English speakers with no Chinese consistently showed a pitch-height-driven response pattern in all the tasks so far using lexical tones.

Non-lexical-tone tasks (Study 3A, 3B, 3C, 3D)

We further investigated cross-modal correspondences using non-lexical tones and IAT tasks. We attempted to compare the results of non-lexical tones and those of lexical tones. We generated the sine waves of Tone 1, Tone 3 and Tone 4 by tracking the pitch and the intensity of some sounds used in our previous studies, with pitch preserved in some of the stimuli in the present study and both pitch and intensity preserved in others. The tone contrasts used here were still the T1-T4 contrast and the T1-T3 contrast.

As seen in our previous studies, when using lexical Mandarin tones, people with different language backgrounds showed different response patterns for tone-angularity audiovisual cross-modal correspondence. However, when using non-lexical tones whose pitch or both pitch and intensity were identical to those in the lexical tone stimuli, all participants (C; C/E; E) showed the same response pattern. That is, all participants matched pointy shapes with high-pitched sounds and curvy shapes with low-pitched sounds for both of the T1-T4 contrast and the T1-T3 contrast without significant difference in the tone-shape mapping effect between subject groups. And the difference in language-dependent effects between speech tones and non-speech tones was significant.

It seemed that pitch, spectral information and other acoustic features interact in Chinese listeners' lexical tone perception. For English listeners, instead of this interaction, pitch height outperforms other acoustic features in their lexical tone perception. While when it comes to non-lexical tones, all participants are more sensitive to pitch height than other acoustic features. As mentioned before, from birth to six months old, infants can distinguish most phonetic contrasts used by the world's languages. However, after about 6 months old, infants start adapting to the sounds they hear most often, coupled with loss of sensitivity to those sounds which are heard infrequently (Kuhl et al., 2006). Our findings that people with different language backgrounds showed different response patterns to lexical tones and showed the same pattern to non-lexical tones suggest that lexical tone perception is acquired and non-lexical tone perception is more basic. However, some studies revealed that language-dependent effects in both speech and nonspeech homologues of lexical tones, e.g., behavioral (Xu, Gandour, & Francis, 2006), electrophysiological responses at the level of the auditory brainstem (Krishnan & Gandour, 2014; Krishnan, Swaminathan, & Gandour, 2009), and functional neuroimaging at the level of the cerebral cortex (Xu, Gandour, Talavage, et al., 2006). Therefore, the difference in language-dependent effects between speech tones and non-speech tones observed in the present study has to be restricted to the sound symbolism task.

6.3 Mappings between visual spatial frequency and Mandarin tones (Study 4A, 4B, 4C)

Besides cross-modal correspondence between visual angularity and Mandarin tones, inspired by the pitch-spatial frequency mapping studies of Evans and Treisman (2010), we investigated the relationship between spatial frequency and Mandarin tones using both lexical tones, non-lexical tones and IAT tasks. The tone contrast in the spatial frequency tasks was the T1-T4 contrast. We detected the ‘*u*-low, *i*-high’ vowel-spatial frequency cross-modal correspondence across the subject groups for the first time, which was novel but not surprising, because the two newly created Gabor patches carried the same spatial frequency information (modes of spatial frequency) as that of the curvy shape and the point shape used in our previous studies. When it comes to tone-spatial frequency mappings, we also found a significant difference in language-dependent effects between speech tones and non-speech tones. When lexical tones were used, as pitch change is important in Chinese listeners’ lexical tone perception, both Chinese groups matched more-pitch-changing tones with high spatial frequency images and less-pitch-changing tones with low spatial frequency images. However, the non-Chinese English-speaking group showed a significantly different mapping pattern from that of the two Chinese groups by pairing high spatial frequency with high pitch and low spatial frequency with low pitch. On the other hand, when non-lexical tones were used, all participants (C; C/E; E) showed the ‘high-high, low-low’ tone-spatial-frequency mapping pattern without significant group difference, which was consistent with the

previously observed cross-modal correspondence between pure tones and spatial frequency in English speakers (Evans and Treisman, 2010). Similar to what was observed in our angularity studies, language specificity existed only when spectral information was involved in cross-modal correspondence between visual dimensions and Mandarin tones.

6.4 Implications

Previous studies have demonstrated certain relationships between visual dimensions (i.e., visual angularity, spatial frequency) and pure tones (i.e., ‘high-pointy, low-curvy’, Marks, 1974; 1987; Walker et al., 2010; ‘high-high, low-low’, Evans and Treisman, 2010). To our best knowledge, our study investigated systematic cross-modal correspondences between visual dimensions (i.e., visual angularity, spatial frequency) and lexical tones (i.e., Mandarin tones) for the first time. We used different approaches including 2AFC tasks and IAT tasks. Using 2AFC tasks, we found that participants with different language backgrounds showed different response patterns. English participants (E) showed a ‘high-pointy, low-curvy’ pattern in terms of pitch height which was consistent with the findings that their lexical tone perception mainly focuses on pitch height (Gandour, 1983) as well as the findings of cross-modal mappings between pure tones and visual angularity of this participant group (i.e., high-pointy, low-curvy, Marks, 1974; 1987; Walker et al., 2010). Chinese dominant bilinguals (C), on the other hand, matched more-changing sounds with pointy shapes and less-changing sounds with curvy shapes, which

was consistent with the focus of lexical tone perception of this language group (i.e., pitch change, Jongman, et al., 2006; Singh, et al., 2014). Somewhat conversely, Chinese-English balanced bilinguals (C/E) swung between the two strategies depending on the tonal contrasts presented. Moreover, the ‘*u*-curvy. *i*-pointy’ vowel-shape mapping has been replicated and has been demonstrated to be stronger than tone effect across participant groups. Language background therefore modulates multisensory perception between visual angularity and lexical tones, but not that between angularity and vowels, as the vowels /i/ and /u/ used here are phonemic in both Mandarin Chinese and English, while only Mandarin speakers are familiar with Mandarin tones not English speakers.

IAT tasks were used in our studies to avoid strategic responses caused by forced-choice tasks and allow for investigation of automatic cross-modal correspondence while only task-relevant stimuli were involved. Previous studies treated Mandarin T1-T3 as the most distinct contrast in terms of pitch height difference (Chandrasekaran, et al., 2007; Singh et al., under review). While in our IAT tasks, especially when lexical tones were used, we found that Chinese dominant participants were more sensitive to T1-T4 contrast (the most contrastive contrast in terms of pitch change), English participants were more sensitive to T1-T3 contrast (the most contrastive contrast in terms of pitch height) and Chinese-English balanced participants were sensitive to both contrasts, and this difference may be due to their balanced bilingual nature. When lexical tones were used, in visual angularity tasks, Chinese dominant participants’ response pattern was based on pitch change (i.e., more change-

pointy, less changing-curvy) especially when vowel and shape did not match, while English participants' pattern relied on pitch height (high-pointy, low-curvy) especially when vowel and shape went with each other. Chinese-English balanced bilinguals showed Chinese-style response pattern when tones were attended and English-style response pattern when tones were hidden. Moreover, overall, the '*u*-curvy, *i*-pointy' vowel-shape mapping pattern has been replicated in the IAT tasks again and vowel effect was stronger than tone effect.

In spatial frequency tasks where the visual stimuli had the same modal spatial frequency information as that of stimuli in visual angularity tasks, when lexical tones were used, both Chinese groups matched spatial frequency with Mandarin tones in terms of pitch change (i.e., more change-high, less change-low) and '*u*-low, *i*-high' vowel-spatial frequency cross-modal mapping has also been observed. On the other hand, when pitch-matched or both pitch-matched and intensity-matched sine waves were used, all participant groups (C; C/E; E) showed 'high-high, low-low' mapping pattern for tone-spatial frequency cross-modal correspondence, which was consistent with the previous reported relationships between pure tones and spatial frequency (Evans and Treisman, 2010), suggesting that language specificity exists in perception of lexical tones but not in that of non-lexical tones. Spatial frequency has also been found as a more powerful driver of tone effects in audiovisual cross-modal correspondence than visual angularity, may be due to its visual simplicity.

Our findings that people with different language backgrounds showed different response patterns to lexical tones and showed the same pattern to non-lexical tones suggest that lexical tone perception is acquired and non-lexical tone perception is more basic. Language exposure and acquisition can reorganize the perception of linguistic sounds, so they, in turn, may influence cross-modal processing of linguistic sounds.

6.5 Limitations and future directions

There are several important limitations in the present study. First, we included a very limited number of phonemes (i.e., /i/ and /u/). The front, unrounded /i/ and back, rounded /u/ have been well documented as potent drivers of vowel-shape mappings with a ‘*u*-curvy, *i*-pointy’ mapping pattern. Moreover, as this was the first investigation of audiovisual cross-modal correspondence between lexical tones and visual dimensions (i.e., visual angularity; spatial frequency), and thus we decided to choose a ‘safe’ vowel contrast as tone carriers, to rule out the possible confounds on tone-angularity correspondences /tone-spatial frequency correspondences caused by vowels. The limited set of vowels indeed allowed us to observe a neat ‘*u*-curvy, *i*-pointy’ sound-shape mapping pattern as well as distinct mapping patterns of Mandarin tones and shapes in different subject groups. However, future studies could include more phonemes and/or more visual dimensions (e.g., /a/, /i/ and visual size) as well as auditory stimuli with combinations of vowels and consonants.

Also, more visual stimuli options may allow for more detailed elaborations on these cross-modal correspondences.

There were also some limitations regarding the missing participant groups. For example, in the spatial frequency studies, due to the recruitment difficulty and the time limitation, the T1-T3 contrast has not yet been examined. It was not very easy to recruit Chinese-dominant participants and non-Chinese English-speaking participants from among the university community at NTU in Singapore. In addition, people who had already taken part in a study were not allowed to take part in another version, as they may know something about the hypothesis. For the Chinese-dominant group and the non-Chinese group, there simply were not enough people in the local community who had not already taken part. In the future, the T1-T3 contrast should be examined in the tone-spatial frequency sound symbolism context to make the comparisons in this kind of cross-modal correspondence more complete.

Another limitation of the current body of research is that only Mandarin tones have been examined in the present studies. To generalize our hypothesis on cross-modal correspondence between lexical tones and visual dimensions, further studies should try to involve lexical tones from various languages, e.g., Thai tones, Cantonese tones or tones from African languages.

Finally, one important limitation is that only adult participants were involved in the present research. Future studies could consider recruiting preverbal infants, e.g., 4-month-olds from different language exposure

backgrounds to investigate whether the language specificity observed in adults' multisensory perception between lexical tones and visual dimensions is also evident in preverbal infants. Future studies could also consider recruiting infants at about nine months of age to investigate whether the first critical period hypothesis for language acquisition (Kuhl, 2004; Kuhl et al., 2006; Mattock & Burnham, 2006; Mattock et al., 2008) is also true for lexical-tone-related audiovisual cross-modal correspondences.

6.6 General Conclusions

The present study extended the previous literature on tone-related audiovisual cross-modal correspondence by employing lexical tones and comparing the results of lexical tones and non-lexical tones across different language groups. This series of studies is the first to show that Chinese-speaking participants matched Mandarin tones with shapes in terms of pitch change within each tone (i.e., the 'steady-curvy, changing-pointy' mapping pattern) and non-Chinese-speaking participants matched them in terms of pitch height (i.e., the 'low-curvy, high-pointy' mapping pattern). Conversely, all participants responded depending on pitch height when non-lexical tones were used, showing a pattern (i.e., the 'low-curvy, high-pointy' mapping pattern) which is similar to the previously reported pitch-shape mapping for non-linguistic stimuli (e.g., Marks, 1974; 1987; Walker et al., 2010). Similarly, Chinese-speaking participants matched Mandarin tones with abstract spatial frequency in terms of pitch change within each tone (i.e., the 'steady-low,

changing-high' mapping pattern) and non-Chinese-speaking participants matched them in terms of pitch height (i.e., the 'low-low, high-high' mapping pattern). And as before, all participants responded depending on pitch height when non-lexical tones were used showing a pattern (i.e., the 'low-low, high-high' mapping pattern) which is similar to the previously reported mapping between pitch and spatial frequency for non-linguistic stimuli (e.g., Evans & Treisman, 2010). The most important finding here is language specificity exists in lexical-tone-related audiovisual cross-modal correspondence but when non-lexical tones are used, people with different language background perform alike in the multisensory perception tasks. These findings suggest that perception of non-lexical pitch is more basic and perception of lexical tones is acquired. The most parsimonious explanation for this pattern of effects is that language experience drives perceptual reorganization of linguistic sounds and this processing change spills over into cross-modal processing of linguistic sounds.

References

- Arata, M., Imai, M., Kita, S., Thierry, G., & Okada, H. (2010). Perception of sound symbolism in 12 month-old infants: An ERP study. *NEUROSCIENCE RESEARCH*, 68, e300.
- Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex*, 63, 196-205. doi: <http://dx.doi.org/10.1016/j.cortex.2014.08.025>
- Atzet, J., & Gerard, H. B. (1965). A STUDY OF PHONETIC SYMBOLISM AMONG NATIVE NAVAJO SPEAKERS. *Journal Of Personality And Social Psychology*, 1, 524-528.
- Baayen, R. H., & Milin, P. (2010). Analyzing Reaction Times. *International Journal of Psychological Research*, Vol 3, Iss 2, Pp 12-28 (2010)(2), 12.
- Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal Perception of Adult and Child Faces and Voices by Infants, 1263.
- Bankieris, K., & Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition*, 136, 186-195.
- Berlin, B. (2006). The First Congress of Ethnozoological Nomenclature. *The Journal of the Royal Anthropological Institute*, 12, S23-S44.
- Bidelman, G. M., & Chung, W. L. (2015). Tone-language speakers show hemispheric specialization and differential cortical processing of contour and interval cues for pitch. *Neuroscience*, 305, 384-392. doi: 10.1016/j.neuroscience.2015.08.010
- Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage*, 59(1), 663-672. doi: 10.1016/j.neuroimage.2011.06.095
- Blackbill, Y., & Little, K. B. (1957). FACTORS DETERMINING THE GUESSING OF MEANINGS OF FOREIGN WORDS. *Journal of Abnormal & Social Psychology*, 54(3), 312.
- Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, 39, 585-594. doi: 10.1016/j.wocn.2011.06.002
- Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, 126, 165-172. doi: 10.1016/j.cognition.2012.09.007
- BROSELOW, E., R. HURTIG & C. RINGEN (1987). The perception of second language prosody. In G. I. S. Weinberger (Ed.), *Interlanguage phonology* (pp. 350-361). Rowley: Newbury House.
- Brown, R. W., Black, A. H., & Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *The Journal of Abnormal and Social Psychology*, 50(3), 388-393. doi: 10.1037/h0046820
- Chan, K. M. (1996). Some thoughts on the typology of sound symbolism and the Chinese language. *Proceedings of the Eighth North American Conference on Chinese Linguistics*, 2, 1-15.

- Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: a multidimensional scaling analysis of the mismatch negativity. *Restor Neurol Neurosci*, 25(3-4), 195-210.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Research Report: Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, 1128, 148-156. doi: 10.1016/j.brainres.2006.10.064
- Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *J Acoust Soc Am*, 128(1), 456-465. doi: 10.1121/1.3445785
- Chao, Y. R. (1948). *Mandarin primer : an intensive course in spoken Chinese*: Cambridge : Harvard University Press, 1948.
- Chuang, C. K., & Hiki, S. (1972). Acoustical Features and Perceptual Cues of the Four Tones of Standard Colloquial Chinese. *Journal of the Acoustical Society of America*, 52(1A), 146.
- D'Onofrio, A. (2013). Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the Bouba-Kiki Paradigm. *Language and Speech*. doi: 10.1177/0023830913507694
- Davis, R. (1961). THE FITNESS OF NAMES TO DRAWINGS. A CROSS-CULTURAL STUDY IN TANGANYIKA. *British Journal of Psychology*, 52(3), 259-268. doi: 10.1111/j.2044-8295.1961.tb00788.x
- Deroy, O., & Spence, C. (2013). Are we all born synaesthetic? Examining the neonatal synaesthesia hypothesis. *Neuroscience and Biobehavioral Reviews*, 37(7), 1240-1253. doi: 10.1016/j.neubiorev.2013.04.001
- Dingemanse, M. (2012). Advances in the Cross-Linguistic Study of Ideophones. *Language & Linguistics Compass*, 6(10), 654-672. doi: 10.1002/lnc3.361
- Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science*, 25(6), 1256-1261. doi: 10.1177/0956797614528521
- Drijvers, L., Zaadnoordijk, L. S., & Dingemanse, M. (2015). Sound-symbolism is disrupted in dyslexia: Implications for the role of cross-modal abstraction processes. Noelle, D.; Dale, R.; Warlaumont, D.S. (ed.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (CogSci 2015)*, 602 - 607. doi: urn:nbn:nl:ui:22-2066/145471
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, 10(1).
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. doi: 10.3758/bf03193146
- Fernández-Prieto, I., Navarra, J., & Pons, F. (2015). How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behavior & Development*, 38, 77-81. doi: 10.1016/j.infbeh.2014.12.008
- Fort, M., Martin, A., & Peperkamp, S. (2015). Consonants are More Important than Vowels in the Bouba-kiki Effect. *Language and Speech*, 58(2), 247-266. doi: 10.1177/0023830914534951

- Fort, M., Weiss A., Martin A., & Peperkamp S. (2013). *Looking for the bouba-kiki effect in prelexical infants*. Paper presented at the 12th International Conference on Auditory-Visual Speech Processing, Annecy, France.
- 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. doi: 10.1016/j.wocn.2007.06.005
- Fromkin, V. A. (1978). *Tone: A linguistic survey*. New York: Academic Press.
- Gallace, A., & Spence, C. (2006). Multisensory Synesthetic Interactions in the Speeded Classification of Visual Size. *Perception & Psychophysics*, 68(7), 1191-1203. doi: 10.3758/BF03193720
- Gandour, J. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 41-76). New York: Academic Press.
- Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11(2), 149-175.
- Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A Crosslinguistic PET Study of Tone Perception. *Journal of Cognitive Neuroscience*, 12(1), 207-222. doi: 10.1162/089892900561841
- Gandour, J. T. K., A. (2014). Neural bases of lexical tone. In H. W. P. Padakannaya (Ed.), *Handbook of South and Southeast Asian psycholinguistics* (pp. 339-349). Cambridge, UK: Cambridge University Press.
- Gandour, J. T. K., A. (2016). Processing tone languages. In G. H. S. L. Small (Ed.), *Neurobiology of language* (pp. 1095-1107). New York: Academic Press.
- Geurts, H. M., Grasman, R. P., Verte, S., Oosterlaan, J., Roeyers, H., van Kammen, S. M., & Sergeant, J. A. (2008). Intra-individual variability in ADHD, autism spectrum disorders and Tourette's syndrome. *Neuropsychologia*, 46(13), 3030-3041. doi: 10.1016/j.neuropsychologia.2008.06.013
- Green, K. P., & Kuhl, P. K. (1989). The role of visual information in the processing of. *Perception & Psychophysics*, 45(1), 34-42. doi: 10.3758/bf03208030
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J Pers Soc Psychol*, 74(6), 1464-1480.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal Of Personality And Social Psychology*, 85(2), 197-216.
- 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, 395-421. doi: 10.1016/S0095-4470(03)00016-0
- Haryu, E., & Kajikawa, S. (2012). Are higher-frequency sounds brighter in color and smaller in size? Auditory-visual correspondences in 10-month-old infants. *Infant Behavior and Development*, 35, 727-732. doi: 10.1016/j.infbeh.2012.07.015
- Haynie, H., Bower, C., Epps, P., Hill, J., & McConvell, P. (2014). Wanderwörter in languages of the Americas and Australia. *Ampersand*, 1, 1-18. doi: <http://dx.doi.org/10.1016/j.amper.2014.10.001>
- Heck, R. H., Thomas, S., & Tabata, L. N. (2012). *Multilevel modeling of categorical outcomes using IBM SPSS*: New York ; London : Routledge, 2012.

Hinton, L., Nichols, J., & Ohala, J. J. (2006). *Sound symbolism*: Cambridge ; New York : Cambridge University Press, 2006.

Digitally printed 1st pbk. version.

Ho, C. S.-H., & Bryant, P. (1997). Development of Phonological Awareness of Chinese Children in Hong Kong. *Journal of Psycholinguistic Research*, 26(1), 109-126.

Hoffmann, R. F. (1978). Developmental Changes in Human Infant Visual-evoked Potentials to Patterned Stimuli Recorded at Different Scalp Locations. *Child Development*, 49(1), 110-118. doi: 10.1111/1467-8624.ep10401411

Holland, M. K., & Wertheimer, M. (1964). SOME PHYSIOGNOMIC ASPECTS OF NAMING, OR, MALUMA AND TAKETE REVISITED. *Perceptual & Motor Skills*, 19(1), 111.

Huang, Y.-H., Pratoomraj, S., & Johnson, R. C. (1969). Universal magnitude symbolism. *Journal of Verbal Learning & Verbal Behavior*, 8(1), 155-156. doi: 10.1016/S0022-5371(69)80028-9

Hwang Gu, S.-L., Gau, S. S.-F., Tzang, S.-W., & Hsu, W.-Y. (2013). The ex-Gaussian distribution of reaction times in adolescents with attention-deficit/hyperactivity disorder. *Research in Developmental Disabilities*, 34(11), 3709-3719. doi: <http://doi.org/10.1016/j.ridd.2013.07.025>

Imai, M., Haryu, E., & Okada, H. (2005). Mapping Novel Nouns and Verbs onto Dynamic Action Events: Are Verb Meanings Easier to Learn than Noun Meanings for Japanese Children?, 340.

Imai, M., Haryu, E., Okada, H., Lianjing, L., & Shigematsu, J. (2006). Revisiting the Noun-Verb Debate: A Cross-Linguistic Comparison of Novel Noun and Verb Learning in English-, Japanese-, and Chinese-Speaking Children. In K. A. H.-P. a. R. M. Golinkoff (Ed.), *Action Meets Word: How children learn verbs* (pp. 450-476): Oxford University Press.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions Of The Royal Society Of London. Series B, Biological Sciences*, 369(1651), 20130298-20130298. doi: 10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109, 54-65. doi: 10.1016/j.cognition.2008.07.015

Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *J Exp Anal Behav*, 85(1), 107-124.

Jespersen, O. (1922). *Language: its nature, development and origin*: New York: Henry Holt.

Johnson, R. C. (1967). MAGNITUDE SYMBOLISM OF ENGLISH WORDS. *Journal of Verbal Learning & Verbal Behavior*, 6(4), 508-511. doi: 10.1016/S0022-5371(67)80008-2

Jongman A., W. Y., Moore C. B., Sereno J. A. (2006). Perception and production of Mandarin Chinese tones. In T. L. Li P., Bates E., Tzeng J. L (Ed.), *Handbook of East Asian Psycholinguistics (Vol. 1: Chinese)* (pp. 209–217). UK: Cambridge University Press.

- Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: An event-related potentials training study. *BMC Neuroscience*, 9. doi: 10.1186/1471-2202-9-53
- Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese Sound-Symbolism Facilitates Word Learning in English-Speaking Children. *Cognitive Science*, 35(3), 575-586. doi: 10.1111/j.1551-6709.2010.01169.x
- Karwoski, T. F., Odbert, H. S., & Osgood, C. E. (1942). Studies in synesthetic thinking: II. The rôle of form in visual responses to music. *Journal of General Psychology*, 26, 199-222.
- Keppel, G., Saufley, W. H., & Tokunaga, H. (1992). *Introduction to Design and Analysis: A Student's Handbook*: Worth Publishers.
- Klank, L. J., Huang, Y.-h., & Johnson, R. C. (1971). Determinants of success in matching word pairs in tests of phonetic symbolism. *Journal of Verbal Learning & Verbal Behavior*, 10(2), 140-148. doi: 10.1016/S0022-5371(71)80005-1
- Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, 114, 19-28. doi: 10.1016/j.cognition.2009.08.016
- Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain Lang*, 110(3), 135-148. doi: 10.1016/j.bandl.2009.03.005
- Krishnan, A., & Gandour, J. T. (2014). LANGUAGE EXPERIENCE SHAPES PROCESSING OF PITCH RELEVANT INFORMATION IN THE HUMAN BRAINSTEM AND AUDITORY CORTEX: ELECTROPHYSIOLOGICAL EVIDENCE. *Acoust Aust*, 42(3), 166-178.
- Krishnan, A., Gandour, J. T., & Suresh, C. H. (2015). Pitch processing of dynamic lexical tones in the auditory cortex is influenced by sensory and extrasensory processes. *Eur J Neurosci*, 41(11), 1496-1504. doi: 10.1111/ejn.12903
- Krishnan, A., Gandour, J. T., & Suresh, C. H. (2016). Language-experience plasticity in neural representation of changes in pitch salience. *Brain Res*, 1637, 102-117. doi: 10.1016/j.brainres.2016.02.021
- Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J Cogn Neurosci*, 21(6), 1092-1105. doi: 10.1162/jocn.2009.21077
- Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res*, 25(1), 161-168. doi: 10.1016/j.cogbrainres.2005.05.004
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat Rev Neurosci*, 5(11), 831-843.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., . . . Lacerda, F. (1997). Cross-Language Analysis of Phonetic Units in Language Addressed to Infants, 684.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age, 606.
- Lapolla, R. J. (1994). An experimental investigation into phonetic symbolism as it relates to Mandarin Chinese. In J. N. J. O. L. Hinton (Ed.), *Sound Symbolism* (pp. 130-147). Cambridge: Cambridge University

Press.

- Leather. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. L. a. A. James (Ed.), *New Sounds 90* (pp. 72–97): University of Amsterdam.
- Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Percept Psychophys*, 53(2), 157-165.
- Leth-Steensen, C., Elbaz, Z. K., & Douglas, V. I. (2000). Mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. *Acta Psychol (Amst)*, 104(2), 167-190.
- Li, C. N. a. T., Sandra A. (1977). *The causative in Wappo: a special case of doubling*. Paper presented at the Annual Meeting of the Berkeley Linguistics Society.
- Lin, C. Y., Wang, M., & Shu, H. (2013). The Processing of Lexical Tones by Young Chinese Children. *Journal of Child Language*, 40(4), 885-899.
- Liu, L., & Kager, R. (2016). Perception of tones by bilingual infants learning non-tone languages. *Bilingualism: Language and Cognition*. doi: 10.1017/S1366728916000183
- Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: a review of behavioural, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.01246
- Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-Symbolism Boosts Novel Word Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Lockwood, G., Hagoort, P. & Dingemanse, M. (2016). How Iconicity Helps People Learn New Words: Neural Correlates and Individual Differences in Sound-Symbolic Bootstrapping. *Collabra*, 2(1), 1-15.
- Lockwood, G., & Tuomainen, J. (2015). Ideophones in Japanese modulate the P2 and late positive complex responses. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00933
- Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*: OUP USA.
- Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans, 20661.
- Maison, D., Greenwald, A. G., & Bruin, R. (2001). The implicit association test as a measure of implicit consumer attitudes *Polish Psychological Bulletin*, 32(1), 1-9.
- Maltzman, I. M. J., Lloyd; Brooks, Lloyd O. (1956). An investigation of phonetic symbolism. *An investigation of phonetic symbolism*, 53(2), 249-251.
- Marks, L. E. (1974). On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness, 173.
- Marks, L. E. (1978). *The Unity of the Senses: Interrelations Among the Modalities*. New York: Academic Press.
- Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3), 384-394. doi: 10.1037/0096-1523.13.3.384
- Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and Weak, 61.

- Mattock, K., & Burnham, D. (2006). Chinese and English Infants' Tone Perception: Evidence for Perceptual Reorganization. *Infancy*, 10(3), 241-265.
- Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106, 1367-1381. doi: 10.1016/j.cognition.2007.07.002
- Maurer, D. (1993). Neonatal Synesthesia: Implications for the Processing of Speech And Faces. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 109-124). Dordrecht: Springer Netherlands.
- Maurer, D., & Mondloch, C. (2005). Neonatal synesthesia: A re-evaluation. In Robertson, C. L & N. Sagiv (Eds.), *Synesthesia: Perspectives From Cognitive Neuroscience*: Oxford University Press.
- Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The Shape of Boubas: Sound-Shape Correspondences in Toddlers and Adults. *Developmental Science*, 9(3), 316-322.
- Miller, J., Daly, J., Wood, M., Roper, M., & Brooks, A. (1997). Statistical power and its subcomponents — missing and misunderstood concepts in empirical software engineering research. *Information and Software Technology*, 39(4), 285-295. doi: [http://dx.doi.org/10.1016/S0950-5849\(96\)01139-1](http://dx.doi.org/10.1016/S0950-5849(96)01139-1)
- Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1152-1164. doi: 10.1037/a0027747
- Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *COGNITIVE AFFECTIVE & BEHAVIORAL NEUROSCIENCE*, 4(2), 133-136.
- Morton, E. (1994). Sound symbolism and its role in non-human vertebrate communication. In J. N. Leanne Hinton, John J. Ohala (Ed.), *Sound Symbolism* (pp. 348-365). UK: Cambridge University Press.
- Navarro-Pardo, E., Navarro-Prados, A. B., Gamermann, D., & Moret-Tatay, C. (2013). Differences between young and old university students on a lexical decision task: evidence through an ex-Gaussian approach. *J Gen Psychol*, 140(4), 251-268. doi: 10.1080/00221309.2013.817964
- Newman, S. S. (1933). Further Experiments in Phonetic Symbolism, 53.
- Nielsen, A., & Rendall, D. (2011). The Sound of Round: Evaluating the Sound-Symbolic Role of Consonants in the Classic Takete-Maluma Phenomenon. *Canadian Journal of Experimental Psychology*, 65(2), 115-124. doi: 10.1037/a0022268
- Nielsen, A. K. S., & Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*(2), 153. doi: 10.1037/a0030553
- Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J., & Greenwald, A. G. (2014). Understanding and Using the Brief Implicit Association Test: Recommended Scoring Procedures. *PLoS ONE*, 9(12), e110938. doi: 10.1371/journal.pone.0110938
- Nuckolls, J. B. (1999). The Case for Sound Symbolism. *Annual Review of Anthropology*, 28, 225-252.

- Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, 112(1), 181-186.
- O'Boyle, M. W., & Tarte, R. D. (1980). Implications for phonetic symbolism: The relationship between pure tones and geometric figures. *Journal of Psycholinguistic Research*, 9(6), 535-544. doi: 10.1007/BF01068115
- Occelli, V., Esposito, G., Venuti, P., Arduino, G. M., & Zampini, M. (2013). The takete-maluma phenomenon in autism spectrum disorders. *Perception*, 42(2), 233-241. doi: 10.1068/p7357
- Ohala, J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. N. Hinton, Johanna; Ohala, John J (Ed.), *Sound Symbolism* (pp. 325-347). UK: Cambridge University Press.
- Ohtake, Y., & Haryu, E. (2013). Investigation of the process underpinning vowel-size correspondence. *Japanese Psychological Research*, 55(4), 390-399. doi: 10.1111/jpr.12029
- Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, 114, 173-186. doi: 10.1016/j.jecp.2012.05.004
- Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *EXPERIMENTAL BRAIN RESEARCH*, 220(3-4), 319-333. doi: 10.1007/s00221-012-3140-6
- Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, 22(2), 237-247. doi: [http://dx.doi.org/10.1016/S0163-6383\(99\)00003-X](http://dx.doi.org/10.1016/S0163-6383(99)00003-X)
- Patterson, M. L., & Werker, J. F. (2002). Infants' Ability to Match Dynamic Phonetic and Gender Information in the Face and Voice. *Journal of Experimental Child Psychology*, 81(1), 93-115. doi: <http://dx.doi.org/10.1006/jecp.2001.2644>
- Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6(2), 191-196. doi: 10.1111/1467-7687.00271
- Peña, M., Mehler, J., & Nespore, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, 22(11), 1419-1421. doi: 10.1177/0956797611421791
- Prom-on, S., Xu, Y., & Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *J Acoust Soc Am*, 125(1), 405-424. doi: 10.1121/1.3037222
- Ramachandran, V. S. (2004). *A brief tour of human consciousness : from imposter poodles to purple numbers*: New York : Pi Press, c2004.
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. *Journal of Consciousness Studies*, 8(12), 3-34.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychol Bull*, 114(3), 510-532.
- Revill, K. P., Namy, L. L., DeFife, L. C., & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain and Language*, 128(1), 18-24. doi: <http://dx.doi.org/10.1016/j.bandl.2013.11.002>

- Saussure, F. d., Riedlinger, A., Bally, C., & Sechehaye, A. (1966). *Course in general linguistics*: New York : McGraw-Hill, [1966].
- Sedley, D. N. (2003). *Plato's Cratylus*: Cambridge, UK ; New York : Cambridge University Press, 2003.
- Singh, L., & Fu, C. S. L. (2016). A New View of Language Development: The Acquisition of Lexical Tone. *Child Development*(3), 834. doi: 10.1111/cdev.12512
- Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: a cross-linguistic investigation. *Developmental Science*, 17(1), 94-109. doi: 10.1111/desc.12097
- Siok, W. T., & Fletcher, P. (2001). The role of phonological awareness and visual-orthographic skills in Chinese reading acquisition. *Developmental Psychology*, 37(6), 886-899. doi: 10.1037/0012-1649.37.6.886
- Sirkin, R. M. (2006). *Statistics for the Social Sciences*: SAGE Publications.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4), 971-995. doi: 10.3758/s13414-010-0073-7
- Spence, C., & Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *i-Perception*, 3(5), 316-318. doi: 10.1068/i0526ic
- Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*, 22(1), 245-260. doi: <http://dx.doi.org/10.1016/j.concog.2012.12.006>
- Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *J Exp Anal Behav*, 65(3), 643-659. doi: 10.1901/jeab.1996.65-643
- Styles, S. J. (2014). *Psychoacoustic correlates of speech-sound/shape/size mapping in a 121-alternative forced-choice task*. Paper presented at the International Multisensory Research Forum, Amsterdam.
- Sučević, J., Savić, A. M., Popović, M. B., Styles, S. J., & Ković, V. (2015). Balloons and bavoons versus spikes and shikes: ERPs reveal shared neural processes for shape–sound–meaning congruence in words, and shape–sound congruence in pseudowords. *Brain and Language*, 145–146, 11-22. doi: <http://dx.doi.org/10.1016/j.bandl.2015.03.011>
- Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport*, 19(11), 1163-1167. doi: 10.1097/WNR.0b013e3283088d31
- Tarte, R. D. (1974). PHONETIC SYMBOLISM IN ADULT NATIVE SPEAKERS OF CZECH. *Language & Speech*, 17(4), 87-94.
- Tarte, R. D., & Barritt, L. S. (1971). PHONETIC SYMBOLISM IN ADULT NATIVE SPEAKERS OF ENGLISH : THREE STUDIES. *Language & Speech*, 14(2), 158-168.
- Taylor, I. K., & Taylor, M. M. (1962). Phonetic symbolism in four unrelated languages. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 16(4), 344-356. doi: 10.1037/h0083261
- Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology*, 64(12), 2392-2404. doi: 10.1080/17470218.2011.605898
- Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, 23(5), 689-708. doi: 10.1080/01690960701728261

- Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. . (1994). *Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6–12 months: language-specific developmental changes in speech perception abilities*. Paper presented at the International Conference on Spoken Language Processing, Yokohama, Japan.
- Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg (Ed.), *Universals of Human Language, Volume 2: Phonology* (pp. 525-568). Stanford: Stanford University Press.
- Van Zandt, T. (2002). Analysis of Response Time Distributions. In J. Wixted (Ed.), *Stevens' Handbook of Experimental Psychology* (pp. 461-516). San Diego, CA: Academic Press.
- Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*, 33(3), 514-535. doi: [http://dx.doi.org/10.1016/0022-0965\(82\)90063-7](http://dx.doi.org/10.1016/0022-0965(82)90063-7)
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences, 21.
- Walker, P., & Smith, S. (1985). Stroop interference based on the multimodal correlates of haptic size and auditory pitch. *Perception*, 14(6), 729-736. doi: 10.1068/p140729
- Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, 113(2), 1033.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106(6), 3649-3658. doi: 10.1121/1.428217
- Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 42(2), 264-280. doi: 10.1016/S0010-9452(08)70352-6
- Whalen, D. H., & Xu, Y. (1992). Information for Mandarin Tones in the Amplitude Contour and in Brief Segments. *Phonetica: International Journal of Speech Science*, 49, 25-47.
- Whelan, R. (2008). Effective analysis of reaction time data. *Psychological Record*, 58(3), 475-482.
- William, S. Y. W. (1977). Acoustical Studies of Mandarin Vowels and Tones (Book). *Modern Language Journal*, 61(5/6), 301.
- Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and Production of Lexical Tones by 3-Year-Old, Mandarin-Speaking Children. *Journal of Speech, Language & Hearing Research*, 48(5), 1065-1079. doi: 10.1044/1092-4388(2005/074)
- Xu, J., Yu, L., Rowland, B. A., Stanford, T. R., & Stein, B. E. (2012). Incorporating Cross-Modal Statistics in the Development and Maintenance of Multisensory Integration. *The Journal of Neuroscience*, 32(7), 2287-2298. doi: 10.1523/jneurosci.4304-11.2012

- Xu, Y., & Emily Wang, Q. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33(4), 319-337. doi: [http://dx.doi.org/10.1016/S0167-6393\(00\)00063-7](http://dx.doi.org/10.1016/S0167-6393(00)00063-7)
- Xu, Y., Gandour, J., Talavage, T., Wong, D., Dziedzic, M., Tong, Y., . . . Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Hum Brain Mapp*, 27(2), 173-183. doi: 10.1002/hbm.20176
- Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J Acoust Soc Am*, 120(2), 1063-1074.
- 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.
- Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123-139. doi: 10.1016/j.jml.2012.09.004
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Appendices

Appendix 2.1

Table A1. Self-proficiency rating scale

1.	“I can speak easily about most things without thinking about how to say them”
2.	“I can talk about many things, but I often need to think about how to put my words together when I speak”
3.	“I can understand most of what people are saying but I can only think how to reply sometimes”
4.	“I can understand some of what people are saying but I usually can’t think how to reply”
5.	“I have taken lessons and have begun practicing how to say certain kinds of words and sentences”

Appendix 3.1

Table A2. Age (Mean & Range) and gender split for each group in each study

	Age Mean			Age Range			Gender (F)			Sample Size		
	C	C/E	E	C	C/E	E	C	C/E	E	C	C/E	E
2A	23.6	21.1	20.9	22- 26	19- 24	19- 25	9	11	9	10	20	10
2B	23.5	21.6	24.2	21- 29	19- 25	20- 28	6	10	4	10	20	10
2C	22.2	21.6	22.6	19- 27	19- 26	19- 28	6	11	3	10	20	10
2D	22.7	20.8	21.5	20- 27	19- 23	19- 22	9	13	6	10	20	10
3A	N/A	21.3	20.3	N/A	19- 25	19- 22	N/A	16	9	N/A	20	10
3B	N/A	20.3	25	N/A	19- 23	13- 32	N/A	10	4	N/A	20	10
3C	23.5	20.9	25.4	21- 26	19- 25	19- 32	10	6	3	10	20	10
3D	24.1	21.7	22.2	22- 29	20- 25	19- 27	5	7	3	10	20	10
4A	24.9	21.3	21.5	23- 31	19- 25	20- 23	9	15	7	10	20	10
4B	24.6	21.5	20.1	20- 30	19- 25	19- 22	9	13	4	10	20	10
4C	26.6	20.5	23.8	22- 30	19- 22	19- 27	8	16	4	10	20	10

Appendix 4.1

Spectrograms for sinewave stimuli

Spectrogram of the sine wave of T1

Spectrogram of the sine wave of T1 (with intensity)

Spectrogram of the sine wave of T3

Spectrogram of the sine wave of T3 (with intensity)

Spectrogram of the sine wave of T4

Spectrogram of the sine wave of T4 (with intensity)