Versatility of phonemic pitch in affective iconicity and perceptual reorganisation

Zheng, T.

Chapter 1
General introduction

When we listen to speech, our ultimate aim is to extract meaning from the acoustic input. Many studies on speech perception have investigated how listeners differentiate phonemes, the fundamental building blocks of speech, from one another (Mitterer & Cutler, 2006). This process relies on a complex interplay between bottom-up (e.g., sensory input) and top-down (e.g., prior knowledge, context, or cognitive expectation) information (Massaro, 2015).

Among many acoustic cues in speech, fundamental frequency (*f0*), perceived as pitch, stands out due to its universal and multiple functions in communication (Gussenhoven, 2004). Across almost all spoken languages, pitch is used both grammatically and paralinguistically, e.g., to mark distinctions between questions and statements or to convey emotions. Specifically, rising *f0* contour is often used for questions, while falling *f0* contour is often used for statements (see reviews in Cruttenden, 1997; Ohala, 1983, 1994). High-arousing emotions generally correlate with a higher overall pitch level (i.e., a more stable contour at a higher frequency), higher average pitch height, wider pitch range, and steeper pitch slope, in comparison to low-arousing emotions (Bänziger & Scherer, 2005; Laukka et al., 2005; Scherer et al., 2003). While some report a correlation between lower average pitch height and positive valence (Scherer & Oshinsky, 1977), the more prevailing view suggests a trend towards a higher pitch level, higher average pitch height, and wider pitch range (Belyk & Brown, 2014; Kamiloğlu et al., 2020; Laukka et al., 2005).

In tone languages, however, pitch serves an additional function: it distinguishes lexical meanings (Yip, 2002). Standard Chinese (SC) is a prototypical tone language which features four distinctive lexical tones: high-level (Tone 1, T1, or H), mid-rising (Tone 2, T2, or R), low-dipping (Tone 3, T3, or L) and high-falling (Tone 4, T4, or F). Some languages, such as Japanese, are not classified as tone languages but still use pitch in a lexically contrastive way. Japanese employs a lexical pitch-accent system, where pitch distinctions are confined to accented syllables within certain words (Kubozono, 2012; Pike, 1948). As Ota et al. (2018) describe, in Tokyo Japanese, words are categorised as either accented, characterised by a phrase-initial rise (realised as %LH), or unaccented, marked by a lexical falling pitch (realised as H*L).

Pitch, despite its diverse functions in discourse, emotional expression, and lexical meaning, demonstrates remarkable versatility in other contexts. For example, it can convey information through pitch form-meaning iconicity, shaped by its perceptual properties and shared linguistic or social conventions among speakers. In addition, distinct patterns of hemispheric lateralisation across languages are related to the perceptual properties of pitch, and its linguistic function. Building on these specific aspects, this dissertation investigates the affective iconicity of linguistic pitch in lexical tones among adult Standard Chinese speakers and the developmental hemispheric lateralisation of lexical-level pitch processing during perceptual reorganisation in Dutch- and Japanese-learning infants. In both cases, pitch perception results from the interplay of bottom-up acoustic information and top-down cognitive, contextual, and experiential influences. By examining these interactions, this work sheds light on the multifaceted role of pitch in speech perception across different contexts, languages, and developmental stages. The following sections of this dissertation will discuss the roles of pitch in affective iconicity and perceptual reorganisation.

1.1 Iconicity

Although lexical items have conventionally been characterised by the seemingly arbitrary association of sounds with meanings (Hockett, 1958), numerous studies have enlightened a significant sound-meaning resemblance, known as iconicity or sound symbolism. Traditionally, iconicity is used when a specific phonetic feature or sound unit, such as phoneme or syllable, is said to go beyond its linguistic function as a contrastive, nonmeaning-bearing unit, to directly express particular meanings (e.g., Ćwiek et al., 2021; Nuckolls, 1999).

Notably, the concept of iconicity is distinguished from systematicity, which refers to a statistically consistent relationship (i.e., statistical regularities) between arbitrary, potentially language-specific patterns of sounds and groups of words that share those patterns (Dingemanse et al., 2015; Thompson & Do, 2019). As argued by Dingemanse et al. (2015), systematicity is presumed to be language-specific due to phonological regularities that are unique to each language. For instance, the onset cluster *gl-* often appears in English words related to light, such as *glimmer*, *glitter*, *glow*, and *glisten* (Sidhu, 2025). While all phonemes that exhibit such patterns are examples of systematicity, not all of them are iconic (Dingemanse et al., 2015; Sidhu, 2025). Iconicity, in contrast, is fundamentally subjective and relies on interpretative processes (Winter et al., 2023). For example, vowels can be iconically associated with multiple meanings across contexts such as size (Blasi et al., 2016), shape (Godoy & Ananias, 2022), colour (Mok et al., 2019), and gender of the entity referred to by the label containing the specific vowel (Wong & Kang, 2019).

Based on these distinctions, Winter et al. (2023) redefined iconicity as follows:

> *A signal in any medium or modality, such as a word, sign, or gesture, is iconic to the extent that language users produce or perceive it through a sense of resemblance between some aspect of its form and some aspect of its meaning*.

In this dissertation, we adopt this definition but focus specifically on how iconicity is perceived. Here, we define iconicity as: A signal in any modality is iconic to the extent that language users perceive it through a sense of resemblance between some aspect of its form and some aspect of its meaning.

Iconicity is a pervasive phenomenon (Adelman et al., 2018; Aryani et al., 2018; Perniss et al., 2010; Winter et al., 2023; Yap et al., 2014) showing universal tendencies (e.g., Auracher et al., 2011; Y.-C. Chen et al., 2016; Huang et al., 1969; Nikroshkina, 2019) while also exhibiting cultural and cross-linguistic variation (e.g., Winter et al., 2021). A widely studied example is the vowel-shape correspondence, where the vowel /i/ is often associated with sharp shapes, and /u/ with round shapes (Köhler, 1929; Ramachandran & Hubbard, 2001). This phenomenon, known as the bouba–kiki effect, has been consistently observed across diverse languages and writing systems (e.g., German, French, Spanish, Polish, Japanese, Korean, Georgian, Chinese, Zulu, Thai, Turkish, Finnish, etc; Ćwiek et al., 2021), and populations, including remote groups like the Himba of Northern Namibia (Bremner et al., 2013), as well as across developmental stages (e.g., toddlers and adults; Maurer et al., 2006). The effect extends to implicit processing, where congruency between spiky shapes and angular letters leads to faster response times (De Carolis et al., 2018). Moreover, the activation of the prefrontal cortex during implicit association tasks suggests that bouba–kiki correspondences may involve modulation of executive processes (Peiffer-Smadja & Cohen, 2019). This implies that bouba–kiki effects may be subject to top-down modulation, with attention and cognitive control playing a role in resolving perceptual conflicts between mismatching stimuli.

Despite its universality, iconicity is influenced by cultural and perceptual experiences. For instance, cultural differences influence shape–sound associations (e.g., Taiwan vs. US participants; Y.-C. Chen et al., 2016), while phonemes (Fort et al., 2015), acoustic cues (Fort & Schwartz, 2022; Knoeferle et al., 2017), phonetic features (D'Onofrio, 2014), and orthography and phonology (Cuskley et al., 2017) further refine these associations. Additionally, the bouba–kiki effect extends to other domains: vowels in names can influence perceptions of facial roundness (Barton & Halberstadt, 2018), and tactile ratings show that /u/ is linked to positive sensations of the materials (e.g., comfort, smoothness, and warmth), whereas /i/ is associated with negative ones (Sakamoto & Watanabe, 2018). Furthermore, the vowel-shape association persists in production tasks, with participants consistently naming shapes using phonemes that align with perceptual findings. For example, spiky shapes were more frequently named with high/front vowels and voiceless stops, whereas round shapes were more often named with back/rounded vowels and lateral or nasal consonants (Godoy & Ananias, 2022).

Among the world's languages, Standard Chinese (SC) stands out as particularly intriguing in the study of iconicity. SC is a tone language where characters correspond to syllables, each carrying distinct phonetic and semantic information. Some studies have explored the character-iconicity in SC, uncovering links between the visual structure of characters and their conceptual or semantic properties (e.g., Koriat & Levy, 1979; F. Li, 2017; V. P. H. Li, 1986; Luk & Bialystok, 2005; Pan & Schmitt, 1996; Xiao & Treiman, 2012), revealing that a small degree of iconicity has been retained in simple characters.

A more intriguing question arises: can the tonal dimension of SC, specifically its lexical tones, exhibit iconicity? Lexical tones in SC involve pitch variations that distinguish word meanings, and their potential for iconic representation may align with the broader phenomenon of pitch iconicity. This question is explored in detail in the following section.

1.2 Iconicity of pitch

Ohala (1984, 1997) is among the first to illustrate pitch iconicity in lexical tones. He demonstrated, based on observations in three tonal languages: Ewe, Yoruba, and Cantonese, that high tones are often associated with connotations of smallness, diminutiveness, and familiarity, while low tones are linked to largeness. Subsequent studies have confirmed the iconicity of lexical tones in Chinese and Cantonese across a range of dimensions, including size, shape, gender, and the personality traits in storybook character naming (Chang et al., 2021; Lapolla, 1995; Shang & Styles, 2017; X. Wang, 2021; Wong & Kang, 2019). For example, in SC, the T4 and vowel /i/ are associated with pointy shapes, while the T1, T2 and vowels /a/ and /u/ are associated with round shapes (Chang et al., 2021; Shang & Styles, 2017). Moreover, Thompson (2018) reports a distributional bias for high-level tones, although only in onomatopoeic words, in Mandarin, Taiwanese Southern Min, and Hong Kong Cantonese. Collectively, these studies reveal a consistent pattern of pitch iconicity, with the pitch characteristics of lexical tones, particularly pitch height, being associated with diverse sensory and semantic dimensions.

An important point is that the development of lexical tones in languages like Standard Chinese is believed to have occurred after the establishment of segmental features (see the review by Michaud & Sands, 2020 and references therein). For example, it is widely acknowledged that lexical tones are derived from the consonants at both ends of a tone-bearing syllable, with prevocalic consonants determining the pitch height and postvocalic consonants determining the pitch contour (e.g., Haudricourt, 1954; Pittayaporn & Kirby, 2017). Thus, research on the iconicity of lexical tones is typically framed in terms of pitch iconicity, since lexical tones are primarily distinguished by pitch variations.

This leads to the broader concept of pitch iconicity, which has been extensively studied and is frequently discussed in relation to the "frequency code" hypothesis (Gussenhoven, 2016; Ohala, 1984). According to this theory, higher and/or rising *fo* signifies smallness, or is associated with socio-cognitive interpretations such as submission and politeness, while lower and/or falling *fo* conveys largeness, dominance or aggression. These cross-modal associations appear robust across languages, including English (Perlman, Clark, et al., 2015), Polish (Rojczyk, 2011), and Dutch (Stel et al., 2012). A comprehensive list of studies on pitch iconicity is presented in the Appendix A, Table A1.

Research on non-linguistic pitch has revealed consistent associations between perceived pitch height and various sensory and perceptual attributes, too. Higher-pitched sounds are linked to small size, sharp shapes, lighter or brighter stimuli, and upward movement, while lower-pitched sounds are associated with larger size, rounded shapes, darkness, heaviness, and downward movement (see reviews in Deroy & Spence, 2013; Spence, 2011).

The resemblance between form and meaning is not confined to the perception of sensory dimensions (e.g., shape and size), however, it also extends to the perception of basic affective meanings, i.e., emotions. Exploring the relationship between sound and emotion is worthwhile, as emotion helps individuals process emotionally relevant information, guide their actions, and adapt to changing circumstances (Kashima et al., 2020).

Building on the concept of iconicity, we define the relation between form and affective or emotional meaning as affective iconicity: A signal in any modality is iconic to the extent that language users perceive it through a sense of resemblance between pitch and affective meanings. This term describes emotional experiences that language users construct through speech forms (Aryani et al., 2019). A more detailed review of affective iconicity follows in the next section.

1.2.1 Affective iconicity of pitch

Few studies have looked into the relationship between lexical tones and affective or emotional meanings. Yao et al. (2013) analysed the lexical tones and emotional meanings of approximately 100 bisyllabic Chinese words and identified a direct link between tonal sequences and specific types of emotion. They observed a distribution bias, with falling tones more common in words denoting anger and joy, high-level tones in words denoting sadness, and rising and low-dipping tones in words denoting fear. Similarly, Yap et al. (2014) proposed that the direction of tonal pitch contours in Chinese words correlates with the emotional valence of their translation equivalents in English, with rising contours perceived as the most positive and falling contours as the most negative.

While tones may relate to affective meaning, it remains unclear whether lexical tones, with their pitch variations, exhibit affective iconicity in tonal languages like Standard Chinese (SC). This question is important for two reasons. First, as previously reviewed, the relationship between pitch and emotional arousal and valence is well-established. High-arousing emotions are correlated with a higher pitch level, higher average pitch height, wider pitch range, and steeper pitch slope compared to low-arousing emotions (Bänziger & Scherer, 2005; Laukka et al., 2005; Scherer et al., 2003). Similarly, wider pitch ranges and higher average pitch height are associated with positive valence (Belyk & Brown, 2014; Kamiloğlu et al., 2020; Laukka et al., 2005).

Second, tonemes in SC are as crucial as phonemes in expressing lexical meanings. Previous research has shown that phonemes can be affectively iconic across languages, suggesting they convey emotional content through their acoustic properties (Schmidtke et al., 2014). This phenomenon, known as affective iconicity, underscores the essential role of emotion in communication (Adelman et al., 2018; Aryani et al., 2018, 2019; Auracher et al., 2011; Darwin, 1998; Davidson et al., 2009; Nielsen & Dingemanse, 2021; Whissell, 2003). Segmental components, such as consonants and vowels, have been shown to predict arousal (German in Aryani et al., 2018; German in Auracher et al., 2020; Japanese in Kambara & Umemura, 2021) and valence (German, Polish, Dutch, Spanish, and English in Adelman et al., 2018; Japanese and German in Körner & Rummer, 2023; Dutch and Chinese in Louwerse & Qu, 2017; Chinese and English in Yu et al., 2021). What remains unexplored is the affective iconicity of lexical tone.

To explore this question, we adopted a dimensional framework for describing emotions, as it captures the complexity of emotional experiences more effectively than the discrete emotion framework (Adelman et al., 2018; Aryani et al., 2018; Bänziger & Scherer, 2005; Kamiloğlu et al., 2020; Laukka et al., 2005). The discrete emotion framework conceptualises emotions as distinct categories, such as anger, happiness, sadness, fear, and surprise. Each category is thought to represent a universal, biologically hardwired response that is consistent across cultures (Ekman, 1992). While this model has been influential, it has been challenged by recent studies because of its unvalidated neural foundations (Posner et al., 2005).

In contrast, the dimensional approach describes emotions along continuous axes, most commonly arousal and valence (Russell, 1980). Arousal reflects the intensity or activation level of an emotion, ranging from calm or passive to highly excited or energetic. Valence, on the other hand, represents the emotional quality, ranging from negative (e.g., sadness, fear) to positive (e.g., joy, contentment) (Barrett & Russell, 1999). While categorical approaches capture the richness of emotional states, these categories are not rigidly discrete; instead, the dimensional framework describes their underlying structure and the relationships between emotions (Cowen & Keltner, 2017). For example, both anger and excitement are high-arousal emotions, but they differ in valence: anger is negative, whereas excitement is positive. Neuroscientific studies have shown that brain responses to emotions are continuously distributed rather than neatly categorised into discrete emotions (e.g., Lindquist et al., 2012). This view is further supported by studies such as Citron (2014) and Hamann (2012), which demonstrate that emotions involve distributed neural patterns rather than localised, discrete responses. The dimensional approach is particularly well-suited for analysing linguistic features, such as tonemes, where emotional expression often operates on a spectrum rather than within discrete categories. This enables researchers to capture the dynamic interplay between acoustic properties and emotional perception.

In summary, we ask the question **R1**: Does pitch, a fundamental cue for emotional prosody, influence emotional responses in lexical tone languages like Standard Chinese, in a way similar to the influence of segmental phonemes on emotional arousal and valence ratings in non-tonal languages?

1.2.2 Linguistic pitch is predominant over phonemes in affective iconicity

Existing research suggests iconicity in SC phonemes and tonemes across various contexts, including onomatopoeias (tones in Thompson, 2018), size perception (tones and vowels /i, u, a/ in Chang et al., 2021), shape (tones and vowels /i, u, a/ in Chang et al., 2021; tones and vowels /i, u/ in Shang & Styles, 2017), power and gender associations (tones and frontal vowels in Shih et al., 2019), and character traits (tones and consonants /p, t, k/ in X. Wang, 2021). However, the interplay between phonemes and tonemes in iconicity remains unclear.

Fort et al. (2015) explored the respective contributions of consonants and vowels in the bouba–kiki effect in French, finding that consonants play a more significant role than vowels. Specifically, varying vowels had a marginally smaller (or negligible) influence on the shape–sound mapping of VCV (or CVCV) pseudowords compared to consonants. The authors proposed that the bouba–kiki effect is influenced by multiple factors, including acoustic, articulatory, and phonological properties of speech stimuli. Tone languages such as SC offer a valuable context for exploring such interactions due to the complexity of their phonetic features across lexical tones, vowels, and consonants.

This study focuses on the interplay between phonemes and tonemes in affective iconicity. Segmental components, such as consonants and vowels, have been shown to predict arousal (e.g., German in Aryani et al., 2018; Japanese in Kambara & Umemura, 2021) and valence (e.g., German, Polish, Dutch, Spanish, and English in Adelman et al., 2018; Japanese and German in Körner & Rummer, 2023; Dutch and Chinese in Louwerse & Qu, 2017; Chinese and English in Yu et al., 2021). In these studies, the segments /i/, /u/, /t/, and /n/ are the most prevalent segments across languages (Styles & Gawne, 2017), which are also significantly relevant to affective iconicity investigations. For instance, research in Japanese and German demonstrated that /i/ is more frequently associated with positive expressions than /u/, suggesting a potential cross-linguistic consistency in vowel-valence iconicity (Körner & Rummer, 2023). Additionally, in poetic contexts, plosive sounds have been linked to pleasant moods and heightened activation, whereas nasal sounds are often associated with unpleasant moods and lower activation (e.g., in German, Chinese, Russian, and Ukrainian, see Auracher et al., 2011). However, findings by Louwerse and Qu (2017) challenge this trend, as Dutch and Chinese speakers rated nasal-initial words differently in terms of emotional valence—positively for Chinese speakers but negatively for Dutch speakers—underscoring the need for further research.

To investigate the interplay between phonemes and tonemes, we included four segmental phonemes (/i/, /u/, /t/, /n/), chosen for their distinct acoustic and articulatory features and frequent occurrence across languages. It is plausible that tonemes exert a greater influence than phonemes due to their phonetic properties. On the one hand, lexical-prosodic features are considered facilitators in interpreting iconic meanings across languages (Dingemanse et al., 2016). On the other hand, studies on emotional expression consistently show that suprasegmental or prosodic features carry more weight than verbal content in languages such as German and SC (Filippi et al., 2017; Lin et al., 2020, 2021). While our goal is not to equate tonemes with emotional prosody and phonemes with verbal semantics, we propose that tonemes play a crucial role in iconicity due to their shared phonetic feature with emotional prosody, pitch, a highly salient perceptual dimension. An affective sensitivity to pitch in both lexical tones and emotional prosody may indicate shared or transferable mechanisms in emotional processing.

This study therefore addresses the following further questions: **R2**: How do SC phonemes exhibit iconicity compared to other languages? **R3**: How do phonemes and tonemes interact in affective iconicity in SC?

Exploring the affective roles of tonemes and their interaction with phonemes could bridge the gap between tonal and non-tonal languages in affective iconicity research. While phonemes have been extensively studied in non-tonal languages, little is known about how tonemes and phonemes contribute to emotional expression in tonal languages. Understanding this dynamic could reveal whether tonal languages prioritise pitch-based emotional communication, offering new insights into the universality and variability of affective iconicity.

1.2.3 Adaptive significance of linguistic pitch in affective iconicity

Although iconicity is prevalent across languages, its underlying mechanisms remain largely unexplored. Iconicity has long been considered a result of perceptual analogies between form and meaning, where linguistic features mirror the properties of the concepts they represent (Nielsen & Rendall, 2011; Sidhu & Pexman, 2018). For instance, Sidhu and Pexman (2018) outlined several potential mechanisms underlying iconicity, including statistical co-occurrence, shared perceptual and conceptual properties, neural factors related to brain structure or information processing, species-general associations such as the frequency code, and language-specific patterns. From a statistical co-occurrence perspective, phonemes may become linked to specific emotions due to the overlap between the musculature involved in articulation and the one used in emotional expression (e.g., Körner & Rummer, 2023). This suggests that iconicity may emerge from the cognitive capacity to perceive and exploit analogies between linguistic form and emotional meaning.

More recently, Adelman et al. (2018) proposed an emotion-centred perspective on iconicity, arguing that iconicity may have evolved due to its adaptive advantages for language users. Specifically, iconicity might facilitate emotional communication, improving the efficiency and effectiveness of language as a tool for conveying affective states. This perspective highlights the functional role of iconicity in emotional expression and suggests that its evolution may be driven by its utility in social and communicative contexts. Across five Indo-European languages, Adelman et al. (2018) explored whether the individual phonemes of words predict the emotional valence of those words. They found that in English, Spanish, Dutch, German, and Polish, the first phoneme of a word predicted its valence better than subsequent phonemes. Interestingly, the faster the first phoneme is produced, the more negative the word is rated. Adelman et al. (2018) attributed this pattern to the phenomenon of automatic vigilance, which refers to the preferential attention to negative stimuli. Based on these observations, the authors suggested that iconicity might be explained as an adaptation, particularly relevant for understanding affective iconicity.

Nevertheless, the adaptation account of iconicity requires further exploration. For instance, investigations into this adaptation account have primarily focused on emotional valence, a dimension shaped significantly by social and cultural conventions (Citron, 2012). In contrast, emotional arousal, which is more closely tied to physiological responses and is considered a more ancient and innate dimension (Darwin, 1998), has received less attention in this context. Expanding the adaptation account to include arousal could provide a more comprehensive understanding of affective iconicity.

Additionally, the adaptation account has primarily been tested in Indo-European languages, where phonemes alone distinguish lexical meanings. In contrast, the majority of the world's languages, i.e., tonal languages, use both phonemes and lexical tones to convey meaning. Lexical tones, distinct from consonants and vowels, primarily signal lexical distinctions but share prosodic features, such as pitch, with emotional intonation. This raises an important question: can the adaptation account of iconicity observed in phonemes also be extended to explain affective iconicity of lexical tones? Given the shared prosodic foundation in lexical tone and emotional intonation, it is plausible that lexical tones might exhibit similar patterns of affective iconicity.

This study aims to extend the adaptation account by shifting the focus from phonemes to tonemes, from valence to both arousal and valence, and from non-tonal to tonal languages. Specifically, we address the research question (**R4**): Does the adaptation account underlying affective iconicity observed in segmental features also apply to suprasegmental features in tonal languages? By exploring this question, the study seeks to advance our understanding of the interplay between iconicity, emotional communication, and linguistic diversity.

1.3 Methodological approaches and stimuli used in affective iconicity research

In the field of iconicity, empirical approaches to measuring affective iconicity include both data-driven and behavioural methods. Data-driven approaches typically utilise large datasets of affective norms to identify phonological components, as demonstrated in German by Aryani et al. (2018). Behavioural approaches typically involve collecting experimental data from participants using rating tasks or guessing-meaning experiments (see review in McLean et al., 2023; Motamedi et al., 2019; Winter & Perlman, 2021). In rating tasks, participants are asked to evaluate how well a signal (e.g., orthographic words, images, or videos) resembles its referent, often using a Likert scale with 5, 7, or 9 points to capture varying degrees of similarity (Motamedi et al., 2019). While less commonly employed in the study of iconicity, slider scales with finer granularity (e.g., 100 points), commonly used in psycholinguistic research (Warriner et al., 2017), may offer valuable insights for future studies on iconicity. The scaled data (Likert/Slider) from these tasks allow for nuanced judgments, especially for real-word stimuli (e.g., Warriner et al., 2013).

In contrast, forced-choice tasks often present participants with an auditory (most likely artificial) or visual stimulus alongside two or more response options, asking them to select the option that best matches the given stimulus (e.g., Imai et al., 2008). Each of these methods provides unique insights into the study of affective iconicity, facilitating a deeper understanding of how signals relate to their referents.

Studies on iconicity employ both real words and pseudowords (or nonce words) to explore the relationship between linguistic forms and their meanings (Lockwood & Dingemanse, 2015; Motamedi et al., 2019). Real-word stimuli provide insights into how existing phonological or morphological patterns align with perceived meanings in natural languages (e.g., Adelman et al., 2018; Hinojosa et al., 2020; Winter et al., 2017, 2023). In contrast, pseudowords, i.e., artificially constructed non-lexical items, are frequently employed to control for prior semantic knowledge and isolate the effects of phonological features on perceived meaning. For instance, studies have used pseudowords to explore cross-linguistic similarities in sound-symbolic associations or to test how phonological properties influence affective or sensory judgments (e.g., D'Onofrio, 2014; Imai et al., 2008; Köhler, 1947; Monaghan et al., 2012). The combination of real-word and pseudoword stimuli provides complementary insights into the nature and mechanisms of iconicity.

1.4 Interim summary on affective iconicity

In Chapters 2, 3, and 4, we systematically investigate the affective iconicity of linguistic pitch in Standard Chinese (SC), focusing on three key aspects outlined in Section 1.2: the emotional associations of linguistic pitch (for lexical tone), the predominant role of tonemes over phonemic components in affective communication, and the mechanisms underlying these associations. This comprehensive investigation aims to provide an integrated understanding of how affective iconicity functions in SC.

Chapter 2 addresses Research Question 1 (**R1**): Does pitch, a fundamental cue for emotional prosody, influence emotional responses in lexical tone languages like Standard Chinese (which use pitch to distinguish words) in a similar way to the influence of how segmental phonemes influence emotional responses, like arousal and valence, in non-tonal languages? To answer this question, three studies were conducted. Study 1 explored the predictive role of lexical tones on emotional arousal and valence ratings using two existing corpus datasets for written bisyllabic words. Emotional ratings in these datasets were collected using 7- and 9-point Likert scales. However, given the limitations of these corpora, such as uneven distributions of lexical tones, and the advantages of a more dynamic Slider scale (Warriner et al., 2013, 2017), Study 2 implemented an online Slider scale rating task. This task, designed for written bisyllabic words, featured improved stimulus balancing, an enhanced design, and advanced statistical modelling to address these limitations. Studies 1 and 2 offered insights into nuanced emotional judgments of real-word stimuli (e.g., Warriner et al., 2013). Study 3 shifted to the auditory domain to further investigate the effect of tonal pitch characteristics. Using spoken bisyllabic nonce words (devoid of semantic meaning), participants completed a two-alternative forced-choice task. This paradigm, commonly used in studies of unfamiliar stimuli, facilitates clear decision-making without overcomplicating the judgment process (e.g., Monaghan et al., 2012). Data across all three studies were collected from native adult SC speakers, combining preexisting corpus data (Study 1) and newly collected experimental data (Studies 2 and 3).

We hypothesised that the studies would reveal a predictive relationship between lexical tones and emotional arousal and valence ratings, reflecting pitch-iconicity effects akin to those observed in emotional prosody. Specifically, we predicted that tonal pitch features—such as pitch level, average pitch height, pitch range, pitch slope, and pitch contour direction—would significantly influence emotional ratings. Higher overall pitch level, average pitch height, pitch range, and pitch slope were expected to correlate with higher emotional arousal (Bänziger & Scherer, 2005; Scherer et al., 2003; Thompson, 2018). Similarly, higher pitch level, wider pitch range, and upward pitch contour direction were predicted to align with positive valence (Belyk & Brown, 2014; Kamiloğlu et al., 2020; Yap et al., 2014). These predictions are grounded in prior findings on pitch, iconicity, and emotional prosody, which together suggest a robust interplay between tonal pitch features and affective judgments in a lexical tone language.

Chapter 3 addresses Research Questions 2 and 3. **R2**: How do phonemes /i, u, t, n/ predict affective iconicity in Standard Chinese? **R3:** How do phonemes /i, u, t, n/ and lexical tones (T1, T2, T3, and T4) interact in predicting affective iconicity in Standard Chinese? To address these questions, auditory contexts were designed to focus on the perceptual emotional experiences of tokens composed of phonemes and tonemes. The study of this chapter employs meaningless monosyllables within a two-alternative forced-choice task. Monosyllabic stimuli were selected over bisyllabic ones to isolate the effects of segmental and suprasegmental components, offering a more straightforward and more controlled examination of each speech unit's contributions. This choice minimises the complexity introduced by tonal coarticulation or by interactions between tones and phonemes across syllables. Additionally, using monosyllables aligns with prior research on the iconicity of vowels and monosyllables (e.g., Chang et al., 2021; Shang & Styles, 2017; Tarte, 1982). The use of nonsense or nonce stimuli further ensures a clear distinction between the perceptual properties of the stimuli and the linguistic factors of interest, avoiding confounding effects of lexical semantics (Lockwood & Dingemanse, 2015). The 2AFC task was selected as the experimental paradigm to enable robust and reliable comparisons (Motamedi et al., 2019).

We hypothesised that both segmental and suprasegmental components contribute to affective iconicity in Standard Chinese. Specifically, we expected T4 would be associated with higher arousal and negative valence due to its fast pitch variation and downward contour. In contrast, T2 and T1 would be linked to positive valence due to their upward contour and stable high pitch level, respectively, and lower arousal due to their less dynamic contours. We anticipated replicating the correspondence of "/i/–positive" and "/u/–negative". We also aimed to explore the consistency of the proposed associations: "/t/–positive and high-arousing" and "/n/–negative and low-arousing". We intended to investigate how different vowels contribute to arousal ratings. For the interaction among phonemes and tonemes, we hypothesised that lexical tones would dominate phonemes in predicting affective iconicity, given that lexical tones rely on primary perceptual cues, namely pitch variations, which are highly salient in iconicity and emotional processing.

Chapter 4 addresses Research Question 4 (**R4**): Does the adaptation account underlying affective iconicity observed in segmental features also apply to suprasegmental features in tonal languages? To address this question, we conducted corpus analyses using the same three corpora that were used in Chapter 2. Specifically, we analysed whether first-syllable tone in bisyllabic words predicted the emotional arousal and valence ratings of the words better than the second-syllable tone. Furthermore, how lexical tones at both syllables differ regarding their predictability of the arousal and valence ratings.

We hypothesised that lexical tone serves to signal both arousal and valence, thereby contributing to the emotional adaptive value of words in a tone language. Results from ERP studies show that higher-arousal emotions elicit more pronounced EPN and LPC components, compared to neutral stimuli. This suggests not only immediate and enhanced voluntary allocation of attention (EPN) but also more extensive and prolonged processing of emotions (LPC) (Fischler & Bradley, 2006; Herbert et al., 2008; Kissler et al., 2007, 2009; Schacht & Sommer, 2009). Such a time course of the neural processing of emotions may result not only from the initial but also the subsequent phonemes in a word that trigger higher emotional arousal. Hence, we anticipated the effects of tone on affective meanings throughout the word. Specifically, valence is more likely to be predicted by a word's initial lexical tone, whereas arousal by both tones of the entire word, due to the distinctive neural processing characteristics of emotional arousal and valence.

These three chapters together provide an overview of how phonemic pitch plays a role in affective iconicity, shaped by lexical tone's pitch properties, linguistic context, and emotional information processing. In the next section, we move toward the processing of linguistic pitch in another context.

1.5 Perceptual reorganisation

The ability to distinguish speech contrasts is fundamental to speech perception and language acquisition. Speech contrasts, which are differences in sounds that convey distinct word meanings, such as phonemes in non-tonal languages or tones in tonal languages, form the foundation for understanding and producing words (Gage & Baars, 2018). For example, in Japanese, the distinction between short and long vowels, such as *kado* ("corner") and *kaado* ("card"), is crucial for meaning differentiation (Hisagi et al., 2010). Similarly, in Dutch, the vowels /ɪ/ and /i/ distinguish words like *rit* ("ride") and *riet* ("reed") (L. Liu & Kager, 2016).

Adults are remarkably specialised in perceiving the speech contrasts of their native language(s), but this specialisation comes at a cost: a reduced ability to perceive contrasts absent from their linguistic environment. This raises a key question: when and how does this specialisation develop?

Research shows that early experiences begin shaping language processing even before birth (Gervain, 2015, 2018). For example, neonates exposed to English during pregnancy display stronger brain activity when hearing English (familiar) sentences compared to Tagalog (unfamiliar) sentences (May et al., 2011). This early sensitivity lays the groundwork for perceptual reorganisation, a process through which infants in their first year of life transition from being universal listeners, capable of distinguishing a wide variety of speech contrasts across languages, to becoming specialised in those relevant in their native language(s) (Werker & Polka, 1993; Werker & Tees, 1984).

By around six months, infants begin to show increased sensitivity to native contrasts and declining sensitivity to non-native contrasts, a phenomenon known as perceptual narrowing or attunement (see reviews in Kuhl, 2004; Singh et al., 2022; Tsuji & Cristia, 2014). For instance, Japanese-learning infants at 6–8 months can still discriminate the English /r/–/l/ contrast, but this ability declines by the end of the first year due to its lack of relevance in Japanese (Kuhl et al., 2006; Tsushima et al., 1994). Conversely, English-learning infants improve in discriminating these sounds between 6–8 and 10–12 months (Kuhl et al., 2006).

Interestingly, perceptual narrowing is not a uniform process and varies depending on the specific speech contrast and linguistic exposure. For instance, Japanese-learning infants can discriminate Japanese vowel length contrasts (/mana/ vs. /ma:na/) at 4 and 7.5 months, but not at 9.5 months (Sato et al., 2010b). Similarly, Dutch-learning infants fail to distinguish the native /i/–/ɪ/ contrast at 5–6 months but succeed by 11–12 months. On the other hand, Mazuka et al. (2014) found that both 4.5-month-old Japanese-learning infants and 10-month-old Japanese-learning infants were able to discriminate syllables containing non-native German vowels (like /bu:k/-/by:k/ and /bi:k/-/be:k/), suggesting that vowel discrimination is influenced by their specific acoustic and/or articulatory properties. In addition, exposure to native vowel tokens that are acoustically similar to the German vowels may facilitate discrimination of the non-native vowels (Mazuka et al., 2014). Together, these findings suggest that the timing of perceptual reorganisation is contrast-dependent and shaped by language experience (L. Liu & Kager, 2016).

Some studies reveal a U-shaped trajectory in perceptual development, where sensitivity to certain contrasts temporarily diminishes before reemerging as infants develop more refined perceptual and cognitive mechanisms. For instance, de Klerk et al.

(2019) found that Dutch-learning infants at both 6 and 10 months could discriminate the native vowel contrast (/aː/-/eː/) and a non-native (English) vowel contrast (/ɛ/-/æ/), whereas 8-month-olds failed to discriminate the non-native contrast. The authors hypothesise that different age groups rely on distinct strategies, with younger infants leveraging early perceptual abilities and older infants using emerging phonetic categories, while 8-month-olds may be in transition between these mechanisms.

The diverse patterns observed across languages and segmental features suggest that perceptual sensitivity to speech contrasts is shaped by a combination of acoustic and/or articulatory properties, linguistic experience, and evolving perceptual strategies.

Beyond segmental features, many of the world's languages rely on suprasegmental features such as lexical tones, primarily cued by pitch variations, to distinguish meaning (Yip, 2002). Pitch is multifunctional, conveying emotional, grammatical, and lexical information, as demonstrated by Chen (2022) in the context of Standard Chinese. This property of pitch makes it a particularly intriguing domain for studying perceptual reorganisation. The next section will explore how these processes unfold for (tonal) pitch contrasts, offering insights into the diversity and complexity of speech perception development.

1.6 Perceptual reorganisation of pitch

Pitch plays a critical role in language acquisition, serving both universal and language-specific functions due to its varied roles across linguistic systems. For instance, in tonal languages like Standard Chinese (SC), pitch is acquired as a marker of lexical tone, emotion, and discourse functions, while in non-tonal languages like English and Dutch, pitch is not used to distinguish lexical meaning but is instead used for sentence-level information and to mark stressed syllables. This dual nature underscores the universality of pitch sensitivity while highlighting its language-specific acquisition pathways.

Neonates exhibit early sensitivity to pitch. For example, French newborns can discriminate Japanese bisyllabic words differing in pitch contour (descending vs. ascending; such as candy /amé/ and rain /áme/) (Nazzi et al., 1998), and Hungarian infants show mismatch negativity (MMN) responses to pitch interval changes in sinusoidal tones, indicating pitch discrimination similar to adults (Stefanics et al., 2009). However, as infants approach their first birthday, their sensitivity to pitch contrasts becomes increasingly shaped by their native language; that is, it becomes perceptually reorganised.

In tonal language environments like Mandarin, pitch sensitivity remains robust and phonologically tuned for distinguishing lexical tones, as tone is crucial for making lexical distinctions. In contrast, in non-tonal environments like English, sensitivity to lexical pitch distinctions diminishes, while sensitivity to sentence-level intonation pitch cues is maintained. For example, Mattock and Burnham (2006) found that between 6 and 9 months, Chinese infants retained the ability to discriminate lexical tones, while English infants lost this ability for speech tones but retained it for nonspeech tones. Similarly, European Portuguese infants at both 5 and 9 months can discriminate intonational phrases consisting of a single prosodic word, distinguishing statements from yes-no questions (Frota et al., 2014). This divergence reflects how infants learning tonal languages encode pitch as phonologically contrastive, while those in non-tonal environments deprioritise pitch variation as lexically irrelevant.

However, mixed findings complicate the understanding of pitch reorganisation across linguistic contexts. Dutch infants who are not learning Limburgian tones can discriminate these contrasts at both 6 and 12 months, just like Limburgian infants (Ramachers et al., 2018). Similarly, 12-month-old Dutch infants succeeded in discriminating SC tone contrasts (e.g., T2 vs. T4; T2 vs. T3) and their musical counterparts, despite failing to do so at 4 months (A. Chen et al., 2017; A. Chen & Kager, 2016). However, Liu and Kager (2014) observed a U-shaped trajectory in Dutch infants' discrimination of SC T1–T4 contrasts, which was discriminated at 5-6 months, not discriminated around 9 months, and discriminated again at 17–18 months.

Neuroimaging evidence links pitch processing to hemispheric lateralisation, with linguistic pitch primarily processed in the left hemisphere (Gandour et al., 2004; Zatorre et al., 1992) and emotional or intonational pitch in the right hemisphere (Gandour et al., 2004; Kotz et al., 2006; Schirmer & Kotz, 2006). Infants' hemispheric responses to pitch also align with developmental trends. Neonates show right-hemisphere responses to intonation (Arimitsu et al., 2011, 2018), and Japanese infants demonstrate a shift from bilateral to left-lateralised processing for lexical pitch-accent within word stimuli (but not for pure tone stimuli) between 4 and 10 months (Sato et al., 2010a). This transition suggests that while pitch discrimination persists, the underlying mechanisms adapt depending on linguistic experience.

The case of Dutch infants raises questions about how their neural processing of pitch-accent contrasts might differ from Japanese infants, given that pitch is not phonologically relevant in Dutch at the word level. Investigating neural responses in Dutch infants could clarify how acoustic properties, linguistic relevance, and neural plasticity interact in pitch reorganisation. By integrating behavioural and neuroimaging evidence, such investigations can elucidate whether behavioural discrimination correlates with hemispheric lateralisation patterns.

Comparing Dutch and Japanese infants would offer a unique opportunity to disentangle the influences of universal auditory principles and language-specific demands on pitch processing, especially during a period of rapid brain maturation (Leroy et al., 2011). Acoustic salience plays a critical role in shaping pitch perception. Certain contrasts, such as pitch accents, may be more easily discriminated due to their distinctive acoustic features, such as pitch contour shape and dynamic range. However, as infants gain linguistic experience, the relevance of these contrasts in their native language environment begins to exert a top-down influence on perception. For Dutch infants, who lack exposure to tone languages, the decline and re-emergence of pitch discrimination may be tied to reduced attention to non-native contrasts during the reorganisation process. Investigating these mechanisms further can clarify how acoustic properties interact with experience-driven factors to shape pitch perception.

We first explore (**R5**) whether Dutch-learning infants can distinguish lexical pitch contrasts, then explore (**R6**) whether Dutch- and Japanese-learning infants show different patterns of lateralisation while processing pitch contrasts. We designed a cross-linguistic study comparing Dutch- and Japanese-learning infants' neural responses when processing pitch contrasts, using functional near-infrared spectroscopy (fNIRS), building on the work by Sato et al. (2010a). The fNIRS technique is ideal for studying infants due to its non-invasive nature, high temporal and spatial resolution, and suitability for awake participants (Wilcox & Biondi, 2015) and has been extensively used in infant studies (Wilcox & Biondi, 2015; Yücel et al., 2021).

We hypothesised that Dutch infants could discriminate lexical pitch contrasts and that initial pitch processing is driven by general auditory mechanisms and acoustic properties. However, language-specific experience would result in distinct processing patterns between Dutch and Japanese infants. Specifically, we anticipated that younger infants exposed to either Japanese or Dutch would show right-lateralised or bilateral responses to pitch contrasts in both non-speech and speech stimuli, reflecting general auditory processing based on pitch's perceptual properties. As they grow older, lateralisation patterns would diverge, particularly for pitch contrasts in speech stimuli, influenced by their linguistic experience with pitch functions. Japanese (older) infants would exhibit left-lateralised responses for lexical pitch contrasts due to their linguistic relevance, while Dutch (older) infants would process those same pitch contrasts as intonation, and would therefore show more bilateral processing.

1.7 Summary of the contents of this dissertation

This dissertation investigates the affective iconicity of linguistic lexical tonal pitch in Standard Chinese in (Chapters 2–4) and the perceptual reorganisation of linguistic pitch in Japanese pitch accent and Dutch lexical stress (Chapter 5). Chapter 2 investigates whether pitch, a key cue for emotional prosody, influences emotional responses in lexical tone languages like Standard Chinese, similar to how segmental phonemes shape emotional expressions, such as arousal and valence, in non-tonal languages (**R1**). Chapter 3 examines the role of phonemes /i, u, t, n/ in affective iconicity (**R2**) and their interaction with lexical tones (T1, T2, T3, and T4) in shaping these associations (**R3**). Chapter 4 extends the discussion by testing whether the adaptation account of affective iconicity observed in segmental features also applies to suprasegmental features in tonal languages (**R4**). Finally, Chapter 5 shifts focus to perceptual reorganisation, first assessing whether Dutch-learning infants can distinguish lexical stress-related pitch contrasts (**R5**) and then comparing how Dutch- and Japanese-learning infants neurologically differ in their ability to discriminate these contrasts (**R6**).