Speech Communication 88 (2017) 17–38

Contents lists available at ScienceDirect

Speech Communication

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

Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal languages

Si Chen[a,*], Caicai Zhang[a,b], Adam G. McCollum[c], Ratree Wayland[d]

[a] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China
[b] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[c] Department of Linguistics, University of California, San Diego, CA, US
[d] Department of Linguistics, University of Florida, FL, US

ARTICLE INFO

Article history:
Received 2 June 2016
Revised 8 December 2016
Accepted 6 January 2017
Available online 11 January 2017

Keywords:
Phonologisation
Perturbation
Functional data analysis
Growth curve analysis
Underlying pitch targets

ABSTRACT

This study statistically models perturbation effects of consonants on f0 values of the following vowel in order to quantify the differences between phonetic perturbation effects (i.e., phonetic variation) and phonologised perturbation effects (i.e., tone distinctions). We investigated perturbation effects in a non-tonal language, Japanese and a tonal language, Chongming Chinese. Traditional methods of modelling cannot distinguish phonetic and phonologised effects on surface f0 contours, as variation caused by both effects reached statistical significance. We therefore statistically modelled and tested the differences in underlying pitch targets, which successfully distinguished between phonetic and phonologised effects, and is robust to data variability. The methods used in this study can be further applied to examine perturbation effects cross-linguistically and shed light on the development of tones and stages of phonologisation more broadly.

1. Introduction

1.1. Gradient and categorical phenomena in speech

The distinction between phonetics and phonology and their relationship has been a subject of an on-going debate (Chomsky and Halle, 1968; Ohala, 1990; Keating, 1996; Steriade, 2000; Flemming, 2001; Keyser and Stevens, 2001; Arvaniti, 2007; Cohn, 2007; Kingston, 2007; Hyman, 2013). According to a modular view, phonology and phonetics are two distinct components of the grammar of sounds in a language; the former deals with discrete and categorical entities (phonological representations), and the latter deals with continuous and gradient phenomena (phonetic implementation) (Chomsky and Halle, 1968). Hyman (2013: 4) summarizes the characteristics that distinguish phonology and phonetics as the following: categorical vs. gradient, discrete vs. continuous, qualitative vs. quantitative, symbolic vs. physical, digital vs. analog, and syntactic vs. semantic. Under this view, the nature of the representation and the operatives within phonology and phonetics are fundamentally distinct. According to Jakobson and Halle's work (1962), a phonological feature refers to a phonetic (articulatory or auditory) property that serves to distinguish a lexical contrast. Physical differences between sounds such as a release burst is not considered a feature because no language has a phonemic distinction between released and unreleased stops (Steriade, 2000). In other words, standard phonological representations only include a subset of physical properties of sounds containing in a word or a phrase, corresponding roughly to its broad transcription (Flemming, 2001). Non-contrastive, fine-grained phonetic details such as segmental duration, timing, precision, coordination, etc., are assumed to be a consequence of universal principles (Chomsky and Halle, 1968), or are supplied by language-specific phonetic component of grammar (Keating, 1990,1994). In language evolution, phonetic variation is interpreted as phonological processes and phonetically motivated sound change leads to recurrent synchronic sound patterns (Blevins, 2004).

A number of criticisms have been raised against compartmentalizing phonetics and phonology to two separate components of the grammar. For example, Steriade (2000) argues that the distinction is neither productive nor enforceable. It is unproductive because phonological patterns cannot be understood without references to their physical manifestations, and it cannot be "coherently enforced" because multiple physical features are simultaneously required to implement most lexical contrasts; to single one out as the phonological feature necessarily involves some degree of

* Corresponding author.
E-mail addresses: sarah.chen@polyu.edu.hk, qinxi3@gmail.com (S. Chen), caicai.zhang@polyu.edu.hk (C. Zhang), agmccoll@ucsd.edu (A.G. McCollum), ratree@ufl.edu (R. Wayland).

http://dx.doi.org/10.1016/j.specom.2017.01.006
0167-6393/

arbitrariness (Steriade, 2000: 314). Similarly, Flemming (2001:10) points out that since phonological features are phonetically specified, it would appear that "sounds are represented twice in the grammar, once at the coarse level of detail in the phonology, and then again at the finer grain in the phonetics" . In addition, obvious parallels have been noted on so called 'phonological' (e.g. assimilation) and 'phonetic' (e.g. coarticulation) phenomena.

These criticisms are evident in more unified, constraint-based models of phonetics and phonology found in some contemporary phonology (e.g., Steriade, 2000; Flemming, 2001; McCarthy and Prince, 1993; Prince and Smolensky, 1993). In Flemming (2001), both categorical (phonological) and gradient (phonetics) phenomena are derived within the same component of grammar, and are subject to the same set of phonetics (speech production) constraints (e.g., minimize articulatory effort), resulting in their observed similarities. However, more integrated models often derive categorical and gradient phenomena differently. For example, in Optimality Theory (OT, McCarthy and Prince 1993; Prince and Smolensky, 1993), which represents an only partially unified model, categorical, non-phonetic contrast-maintenance faithfulness constraints such as "Don't deviate from inputs, IDENT" interact with phonetic constraints to yield an optimal output. In unified models, like Flemming (2001), there is little to differentiate categorical from gradient, and both are derived entirely from phonetic constraints (see Cohn, 2007 for a more modular account of categorical and gradient phenomena). In both Flemming's model and OT, outputs that optimally satisfy conflicting constraints are selected. However, unlike OT, constraint conflicts are resolved by a weighting system rather than a strict dominance system in Flemming's (2001) model (see also weighted Optimality theoretic grammars, e.g. Pater, 2009).

Regardless of theoretical accounts on how speech is underlyingly represented and derived, it is commonly acknowledged that surface, physical manifestation of speech exhibits a great deal of variation, simultaneously signalling its targets, functional, contrastive units while satisfying articulatory constraints. Thus, the central question in speech perception is (and has been for nearly six decades), how listeners separate 'substance' from 'form' in the physical signals. Broadly, two theoretical approaches have been proposed to provide an answer. The first approach, represented by the Motor Theory, suggests that speaker's intended, invariant neuromotor commands associated with underlying articulatory targets is retrieved from the acoustic signals by a specialized, speech-specific neural network (Liberman and Mattingly, 1985; Liberman et al., 1967). In contrast, the general auditory approach argues that objects of speech perception are auditory or acoustic events present in the speech signals. Relying on the same auditory and cognitive mechanisms evolved to perceive other sounds in the environment, the humans' auditory processing system is sensitive to statistical regularities in the distributions of acoustic properties as they co-vary with phonemic distinctions in different contexts (e.g., Diehl, Lotto and Holt, 2004). Evidence of either intrinsic or extrinsic normalization processes during speech perception lends support to this latter account of speech perception mechanism (e.g., Johnson, 2008; Zhang and Chen, 2016).

The overall goal of this current study is to better understand the differences between categorical and gradient effects on the physical realization of speech. Specifically, the study attempts to show that these two effects may be separated using statistical modelling (see Shih, 2005 for an analysis of Mandarin Tone 2 sandhi), thus allowing underlying functional targets of speech to be directly extracted from its physical, acoustic signals and compared statistically. If successful, such approach would not only provide new insights into the human's speech perception mechanism, but also significantly improve computerized speech recognition systems.

The study has 3 specific aims. The first aim is to explore categorical and gradient perturbation effects in a tone (Chongming Chinese) and a pitch-accent (Japanese) language. The second goal is to determine statistical modelling procedures that can most effectively differentiate gradient and categorical perturbation effects in both languages. The third goal is to compare results of different methods of statistical modelling and to extend the models for future investigations of categorical and gradient pitch (f0) phenomena in the world's languages.

The remainder of the paper is organized as follows. Sections 1.2–1.5 introduce phonetic and phonologised perturbation effects, statistical modelling of surface f0 contours and underlying pitch targets, and present hypothesized situations where modelling procedure may differentiate between the two. Section 2 evaluates all proposed statistical methods with Japanese data. Section 3 further applies them to Chongming Chinese data. Sections 4 and 5 provide discussion and conclusion of the results, as well as their implications for future studies of phonetic and phonological perturbation.

1.2. Phonologisation of perturbation effects

It is generally acknowledged that surface representations of underlying segmental units such as vowels and consonants show a great deal of variation (e.g., Lindblom, 1963; Öhman, 1966; Steven and House, 1963). A similar conclusion has been reached for the physical manifestation of suprasegmental phenomena in speech including tone, pitch accent and intonation, whose main acoustic correlate is f0 level or f0 contour (e.g., Xu and Wang, 2001). A number of articulatory constraints contributing to observed f0 variations have been documented including vowel intrinsic f0 (e.g., Lehiste and Peterson, 1961; Shi and Zhang, 1987; Whalen and Levitt, 1995) and initial consonant f0 perturbation (Hombert, 1978; Howie, 1974; Lehiste, 1975; Lehisted and Peterson, 1961)

Perturbation effects of preceding consonants on f0 are widely noted in many languages, both tonal and non-tonal (Abramson and Lisker, 1985; Gandour, 1974; Hyman, 1973a, 1973b). Some consonants tend to raise f0 and others lower it. For example, fricatives exert a greater f0 raising effect than stops in Mandarin Chinese (Shih, 2001). Moreover, initial voiced consonants exhibit an f0 lowering effect on the following vowels whereas voiceless consonants exert the opposite effect. These effects have been attested in a number of languages including Yoruba (Hombert, 1978), Siamese (Gandour, 1974), Yucatec Maya (Frazier, 2009) and Phuthi (Donnelly, 2009). The effects of voiceless unaspirated vs. voiceless aspirated consonants on pitch are also reported, though not as consistent as the effects of onset voicing (see Chen, 2011 for a summary).

These perturbation effects have been claimed to play a role in the phonologisation of f0 and ultimately the development of lexical tone contrasts. For instance, f0 perturbation caused by consonants formed the basis for the widely adopted theory of tonogenesis (Haudricourt, 1954; Chen, 2000; Hombert et al., 1979; Matisoff, 1973; Rose, 2002; Svantesson and House, 2006). In his analysis of the origin of lexical tones in Vietnamese, Haudricourt (1954) proposes that f0 perturbation effects of initial and final consonants on the following and the preceding vowels play a direct role in the development of the six lexical tones in Vietnamese: proto initials determine pitch height or register (high vs. low) whereas proto final consonants determine pitch contour (level, falling and rising). Thurgood (2002, 2007) replaces Haudricourt's consonant-based account with a laryngeal-based account of tonogenesis, arguing for an intermediary stage of voice quality distinctions (e.g., breathy, clear and creaky) developed after initial proto voiced and voiceless, and proto final voiced sonorants, stops and voiceless fricatives, which are responsible for pitch height and pitch contour distinctions in Vietnamese. In other words, Thurgood argues

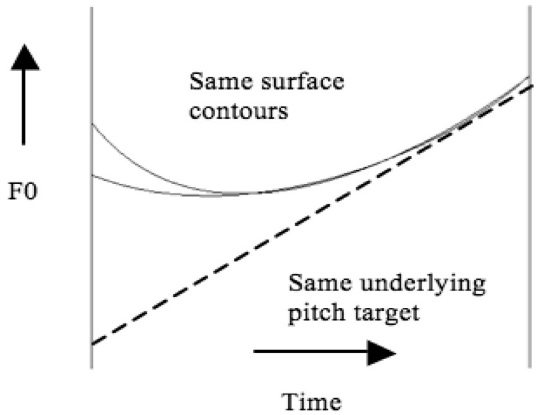

Fig. 1. Small perturbations with no significant differences in surface f0 contours and underlying pitch targets. Surface f0 contours are represented by solid lines and underlying pitch targets are indicated by dotted lines.

that f0 perturbations associated with voice quality or phonation type plays a direct role in tonogenesis. A similar account has been proposed for tonogenesis in a variety of languages including San Martín Itunyoso Trique (DiCanio 2008), Manange (Hildebrandt, 2003), Tamang (Mazaudon and Michaud, 2008, 2012), Middle Chinese (Pulleyblank, 1978) and Wujiang Chinese (Ye, 1983). In addition, Chen (2000) argues that tones may split according to aspiration. Tonal bifurcation from aspiration differences is described in Lengshuijiang Chinese (Zhang, 2009a), Wujiang Chinese (Ye, 1983) and Manange (Hildebrandt, 2003:15).

Tonogenesis from onset voicing, aspiration and phonation are examples of phonologisation processes (Haudricourt, 1954; Huang, 1995; Hyman, 2013: 4; Shen, 2011; Yip, 1980). Within the phonological literature, Yip (1980) incorporates the idea of tonal bifurcation into her phonological analysis of tones. Based on different phonation types, Zhu (2012) also proposes the "Multi-Register and Four-Level tonal model" to improve the traditional "five-point" tonal scale (Chao, 1930).

This study represents the first attempt to distinguish phonetic perturbation patterns on f0 driven by onset consonant voicing or aspiration and phonologised f0 perturbation using statistical modelling procedures.

1.3. Quantitative models of surface f0 contours and underlying pitch targets

We employed the conceptual framework outlined in Xu and Wang (2001) to guide our investigation. According to Xu and Wang (2001), a surface f0 contour, in and of itself, is not the abstract linguistic unit of a lexical tone, but rather its physical realization, which is necessarily subject to articulatory constraints imposed upon speech production. To impart meaning distinctions conveyed by lexical tone, it is necessary to distinguish underlying pitch targets or "the smallest articulatorially operable units associated with linguistically functional pitch units", from surface f0 contours (p. 321). Within this framework, it should be possible to separately model the effects of onset consonants on surface f0 contours and their associated underlying pitch target(s) as exemplified in Figs. 1–3 below.

Figs. 1–3 show three hypothetical levels of consonantal perturbation effects on f0 contours of following vowels. The statistical significance of surface f0 contours is not determined only by the mean values of contours as shown in the figures, but by modelling contours produced by multiple speakers before any further comparison. The same procedure is followed for statistical modelling and comparison of the underlying pitch targets. Fig. 1 depicts the

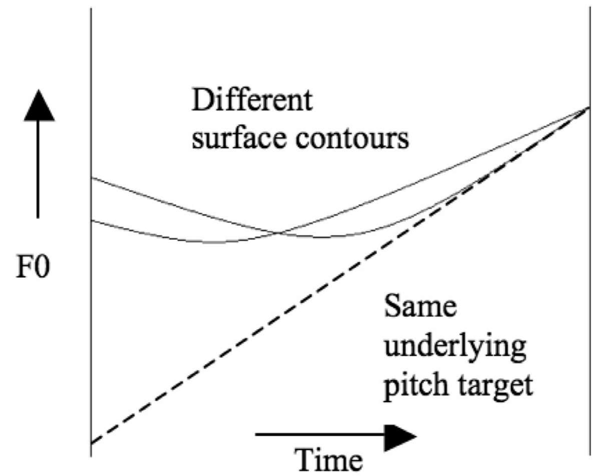

Fig. 2. Significant differences in surface f0 contours not in underlying pitch targets. Surface f0 contours are represented by solid lines and underlying pitch targets are indicated by dotted lines.

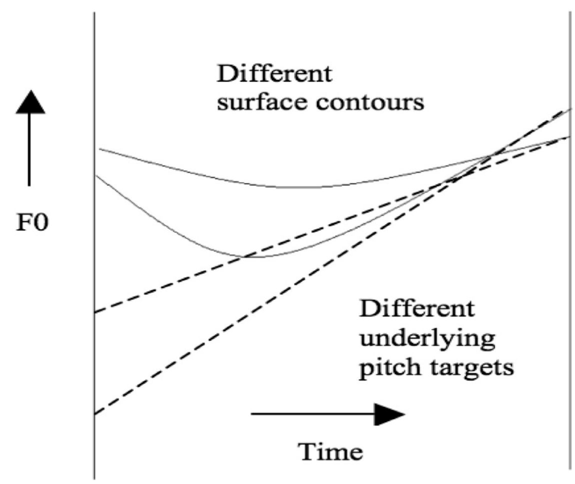

Fig. 3. Significant differences in surface contours and underlying pitch targets. Surface f0 contours are represented by solid lines and underlying pitch targets are indicated by dotted lines.

scenario where two surface f0 contours of the same underlying pitch target are perturbed by different prevocalic consonants, but not to a significantly different degree. In this case, statistical modelling shows a non-significant difference between the two surface contours, and as a result, a non-significant difference between their underlying pitch targets is also expected. On the other hand, statistical modelling of the observed f0 contours in Fig. 2 is expected to reach significance due to a larger degree of initial consonant perturbation on each contour, but without any significant difference in underlying pitch targets. In Fig. 3, the two contours are expected to significantly differ since they represent two different underlying pitch targets. In short, a difference in surface f0 does not imply a necessary difference in underlying pitch targets. In contrast, though, a difference in underlying pitch targets necessarily implies distinct surface f0 contours.

1.3.1. Surface f0 contours modelling

To test the differences between two curves (i.e. two surface f0 contours), many statistical methods are available, including generalized additive model (Hastie and Tibshirani, 1986, 1990; Ning, Shih and Loucks, 2014; Wieling et al., 2014; Wood, 2006), growth curve analysis (Mirman, Dixon and Magnuson, 2008; Mirman, 2014), polynomial regression (Andruski and Costello, 2004; Grabe, Kochanski and Coleman, 2007; Shih and Lu, 2015) and functional data analysis (Gubian et al., 2015; Ramsay and Silverman, 2005). A brief review of each method is offered below.[1]

The generalized additive model (GAM) developed by Hastie and Tibshirani (1990) compares values over a number of time points. One advantage of GAM is that after fitting the model, the role each predictor plays in the model can be examined separately (Chen, 2015a; Wood, 2006). Another advantage of GAM is that it does not assume linear dependence among predictors, which may not exist in the data, in contrast to the standard parametric multiple linear regression model (Hastie and Tibshirani, 1990). However, since GAM is applicable only when multiple independent variables are involved (see Chen (2015a) for incorporating time point and intensity values), it is not suitable for our study since 'time point' is our only independent variable.

Growth curve analysis uses multilevel regression to study time course data. It uses orthogonal polynomials with the advantage of uncorrelated linear and quadratic terms. It is also advantageous in allowing subject-specific deviation of slope over time (Mirman, Dixon and Magnuson, 2008; Mirman, 2014). For our data of f0 values, this method is better than a statistical test based solely on a polynomial regression and does not model variation of individual's trajectories over time.

Functional data analysis is another technique that compares time series data and captures their correlation. It consists of a collection of statistical methods including steps such as smoothing and interpolation of data, data registration or feature alignment, where information about the derivatives of the curves can also be taken into consideration (Ramsay and Silverman, 2005). The advantages of functional data analysis include its usage of continuous smooth dynamics for accurate parameter estimation by creating functional data out of discrete observations over time, resulting in a significant noise reduction due to smoothing techniques (Ullah and Finch, 2013).

In this study, we used growth curve analysis to model surface f0 contours and to test whether these analyses may differentiate phonetic from phonologised perturbation effect. In addition, functional data analysis was used to obtain specific locations where two curves showed statistical differences (Ramsay and Silverman, 2005, 2009), thus providing detailed information about phonetic perturbation, and how it differs across tonal types or pitch accent patterns.

1.3.2. Underlying pitch targets

Xu and Wang (2001: 321) define underlying pitch targets as "the smallest articulatorially operable units associated with linguistically functional pitch units such as tone and pitch accent" (p. 321). A pitch target may be a static one, such as [high] or [low], or a dynamic one, such as [rise] or [fall], and surface f0 contours are viewed as the realization of the underlying pitch targets, or alternatively, as the surface acoustics, generated through target approximation (TA) (Xu, Lee, Prom-on, Liu, 2015). In other words, observed f0 contours are characterized by both functional as well as articulatorially obligatory properties (Xu, 2005). These two properties are modelled separately in the quantitative target approximation model, one as the driving force of a linear system, and the other as sequential target approximation respectively (Prom-On, Xu and Thipakorn, 2009). In this study, we adopt this framework, not only for the purposes of synthesizing f0 contours, but also for statistically testing the differences between underlying pitch targets, which offers the possibility to differentiate phonetic and phonologised perturbation quantitatively.

[1] Previous studies used repeated measures ANOVA to relate f0 values extracted at certain time points to a set of covariates. Although it is among the earliest proposals to deal with correlated responses, and is still widely used [Ma, Mazumdar and Memtsoudis, 2012], this method has been criticized by statisticians [Gibbons, Hedeker and DuToit, 2010], and accordingly, the current study uses more advanced statistical modelling procedures.

1.4. Background on Japanese and Chongming Chinese

Chongming Chinese and Japanese were selected because perturbation with respect to voicing is phonologised as tones in Chongming Chinese, but not in Japanese. In addition, there is a three-way contrast involving voicing and aspiration in Chongming Chinese, which is relatively rare in the majority of Chinese dialects. Japanese, on the other hand, has only a two-way contrast in onsets: voiceless vs. voiced.

Japanese is a pitch-accent language. Similar to lexical tone, pitch accent is used to mark lexical contrast and its location in a word is unpredictable. However, unlike in lexical tone languages where each syllable is unpredictably associated with a tone, a tone on each mora in Japanese is predictable by rules once the accent location, defined as the point at which the pitch falls, is known. More importantly, the onset consonant does not affect the assignment of accent (Bennett, 1981; Haraguchi, 1977; McCawley, 1977). Kubozono (2011) notes that accent usually marks phonologically prominent positions in a word. There are two pitch accent patterns in Japanese: L-H and H-L.

Some early studies examining perturbation effects in Japanese show consistent results that voiced onsets are associated with lower f0 values, and voiceless onsets are associated with higher f0 values. Shimizu (1989,1994) reports lower mean f0 values after voiced onsets. Moreover, Shimizu (1989) observes rising f0 curves for 60 ms after vowel onset following voiced stops. Kawasaki (1983) identifies a 40 ms dip in f0 after voiced stops, and in the H-L accent, f0 peaks later after voiced stops, but in the L-H accent, f0 increases rapidly after an initial dip. Ishihara (1998) shows that although voiced onset consonants are sometimes devoiced, the perturbation effects observed after voiced and voiceless onset consonant are consistent. Using a mixed effects logistic regression model, Kong (2009) also shows that f0 values on a Japanese female speaker's tokens contribute significantly to the discrimination of voiced and voiceless stops. Fig. 4 plots averaged f0 contours from all recorded speakers in this study on Japanese monosyllables, disyllables with L-H accent and H-L accent after voiced vs. voiceless onsets.

Chongming Chinese, also called Haimen, Qidong or Qihai Chinese, is a northern Wu dialect spoken primarily in Chongming County of eastern China. Chongming Chinese is spoken not only in Chongming County, as well as in Haimen, Qidong City, Shazhou County and other areas such as Nanhui, Fengxian and Chuansha. While there is very little phonological variation reported within these regions (Zhang, 2009b), it is suggested that the tone system and tone sandhi behaviour of Chongming Chinese may differ between younger and older speakers (Zhang, 2009a). To control for intergenerational differences, the current study focuses on speech productions of older speakers.

Chongming Chinese has eight contrastive tones (Zhang, 2009). Using Chao's (1930) 5-point pitch scale, where 5 represents the highest pitch value, and 1 represents the lowest, the eight tones in Chongming and their corresponding Middle Chinese tones are described in Table 1 (Chen and Zhang, 1997). There were four tonal categories in Middle Chinese (from around 200 to 900 A. D.): ping "level", shang "rising", qu "departing" and ru "entering", as recorded in the Lu Fa-yan's rhyme book *Qieyun* (AD601) (Chen, 2000; Lu, 2013). Zhongyuan Yinyun (Zhou, 1324) indicates that Middle Chinese level and rising tones have split into Yin and Yang (Xu and Fu, 2015). Wang (1967) and Cheng and Wang (1977) also argue for a register split of each tone category into two contrasting tones, either the Yin (high) subgroup conditioned by voiceless initials or the Yang (low) subgroup conditioned by voiced initials.

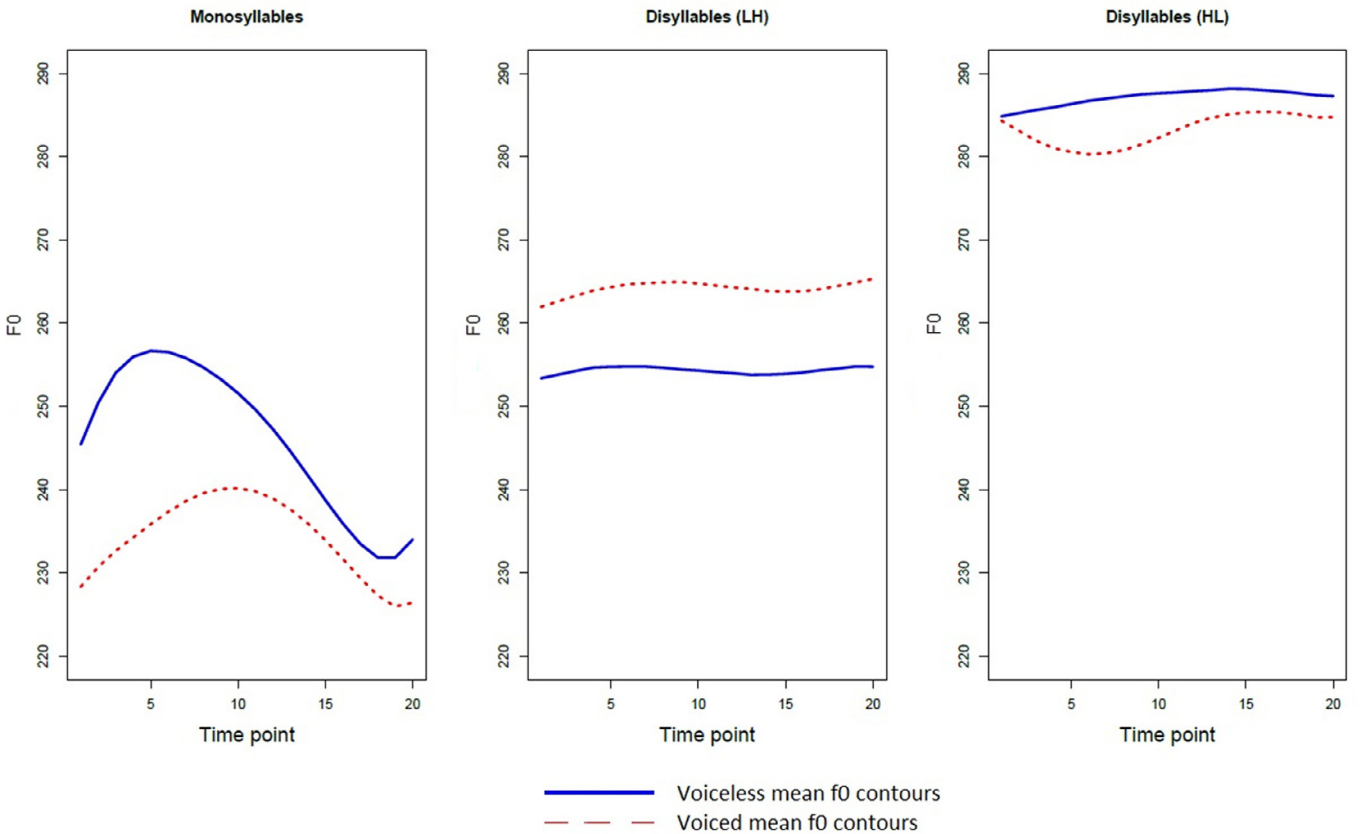

Fig. 4. Averaged f0 contours on Japanese monosyllables, disyllables with L-H accent and H-L accent. The solid blue lines stand for mean f0 contours after voiceless onsets, and the dotted red lines stand for mean f0 contours after voiced onsets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Eight tones in Chongming Chinese.

Middle Chinese categories		Ping (Level) Even	Shang (Rising) Oblique	Qu (Departing) Oblique	Ru (Entering) Oblique
Chongming tones	High register	Tone 1 H (53)	Tone 3 HMH (435/424)	Tone 5 M (33)	Tone 7 Hʔ (55/5)
	Low register	Tone 2 LM (24)	Tone 4 LML (241/242)	Tone 6 MLM (213/313)	Tone 8 Lʔ (23/2)

Significantly, the voicing contrast of Middle Chinese is lost in most modern Chinese dialects (Pulleyblank, 1991; Chen, 2000). Chongming tones are divided into four high-low register pairs: T1 (53) - T2 (24), T3 (435/424) - T4 (241/242), T5 (33) - T6 (213/313) and T7 (55/5) - T8 (23/2) as shown in Table 1 (Chen and Zhang, 1997).

Chongming Chinese is reported to have a three-way contrast among onset obstruents (voiceless aspirated, voiceless unaspirated and voiced) (Chen and Zhang, 1997). Fieldwork records show phonologisation of tones due to the effects of voicing but not aspiration (Chen and Zhang, 1997). An examination of a Chongming Chinese dictionary (Li and Zhang, 1993) and the data collected by Zhang (2009) reveal complementary distribution for the high- and low-register tones in each pair. In T1 (53), T3 (435/424), T5 (33) and T7 (55/5), the carrier syllables have voiceless aspirated or unaspirated obstruent onsets, while in T2 (24), T4 (241/242), T6 (213/313), T8 (23/2), the carrier syllables have a voiced onset obstruent. Fig. 5 plots averaged f0 contours from all recorded speakers on Chongming Chinese monosyllables after three types of onsets.

It is generally agreed that the four tonal pairs developed into eight tones via bifurcation after voiced vs. voiceless onsets (Mei, 1970; Ting, 1996; Chen, 2000). An anonymous reviewer pointed out that an alternative phonological analysis can also be that Chongming Chinese only has four tones phonologically, where the high vs. low registers are phonetic in nature, considering the existence of voicing contrasts in onset obstruents. The traditional eight-tone analysis is more consistent with the tonal analysis across Chinese dialects (e.g. Chen, 2000). Thus, we adopt this generally-agreed view, and model eight rather than four underlying tone targets before voiced and voiceless initials. After phonologisation of these tonal distinctions, surface contours in Chongming Chinese may have changed considerably.

1.5. The current study

This study focuses on statistical modelling of perturbation effects on surface f0 contours and underlying pitch targets following different types of onset consonants in Japanese and Chongming Chinese. Japanese obstruents exhibit a two-way contrast: voiced vs. voiceless onsets (e.g. Shimizu, 1989). On the other hand, Chongming Chinese has a three-way contrast for obstruents: voiced, voiceless aspirated and voiceless unaspirated consonants (e.g. Chen and Zhang, 1997).

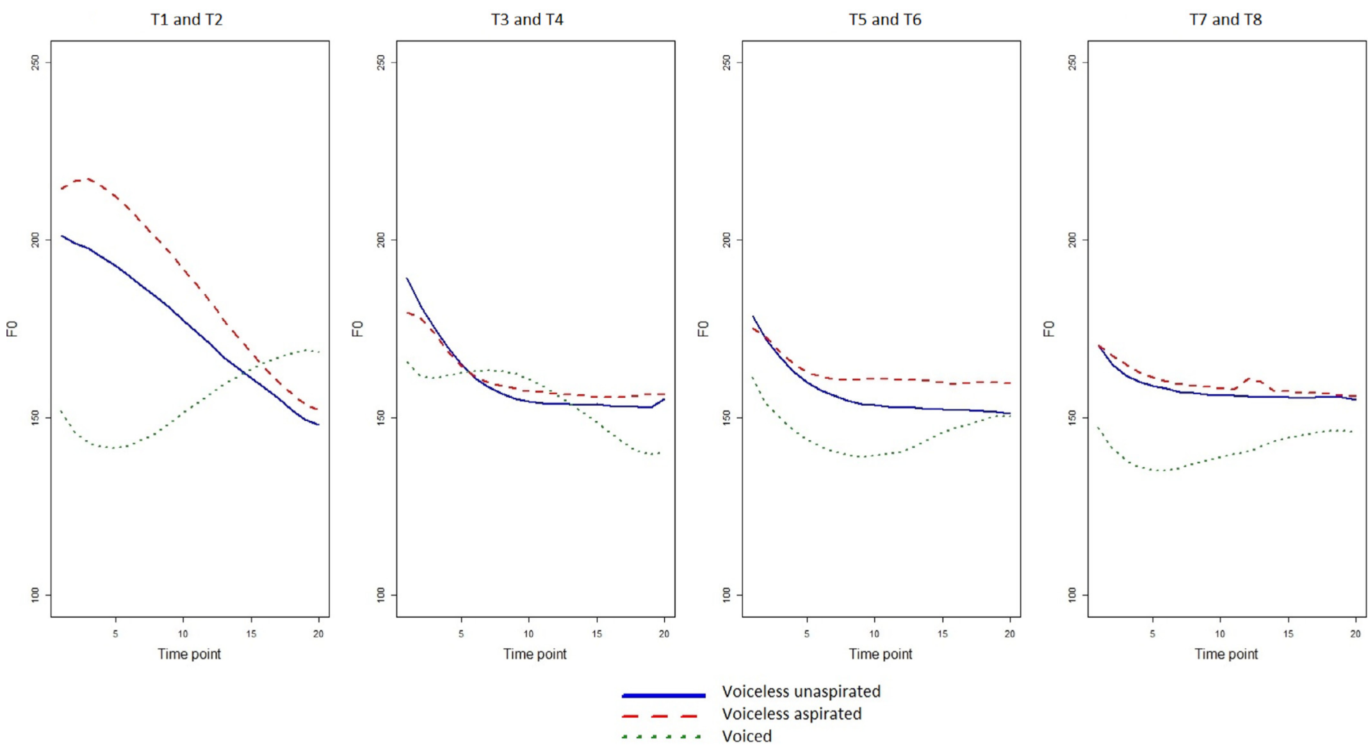

Fig. 5. Averaged f0 contours from all recorded speakers on Chongming Chinese monosyllables after three types of onsets. The solid blue lines stand for mean f0 contours after voiceless unaspirated onsets. The dashed red lines stand for mean f0 contours after voiceless aspirated onsets. The dotted green lines stand for mean f0 contours after voiced onsets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Following the framework of Xu (2005), we assume that observed f0 contours are articulatorially-constrained realizations of underlying pitch targets. We further assume that phonologisation of tones is associated with greater differentiation of f0 than mere phonetic perturbation. Using Hyman's (1976) proposal that phonologised effects are exaggerated phonetic effects (see also Przezdziecki (2005) for vowel harmony), statistical methods were applied to differentiate between phonetic and phonologised (exaggerated) perturbation effects. Similar to hypothetical scenario in Figs. 1–3 in Section 1.3, we tested whether modelling surface f0 contours and underlying pitch targets can differentiate the two types of perturbation effects.

2. Acoustic analysis of Japanese

2.1. Subjects and materials

Thirteen native speakers of Japanese (three males, ten females) between the age of 19 to 50 years old were recruited from Hong Kong Polytechnic University and Beijing Normal University (Zhuhai campus). No participants reported any history of speaking, hearing or language impairment. None of the participants reported experience learning tonal languages, though they likely had at least some exposure to Cantonese and/or Mandarin, since they were living in China at the time of recording. Following Kawasaki (1983) and Ishihara (1998), we collected monosyllables and disyllables bearing low-high (L-H) and high-low (H-L) accent patterns. Each participant read CV monosyllables embedded in a carrier sentence "ima CV wo itte kudasai (Please read CV now)" three times respectively, where C=[p, t, k, b, d, ɡ], and V=[a, e, o]. Target words were embedded in the carrier phrase to control for phrase-level effects on f0. Disyllables bearing different accent patterns (L-H and H-L) were also recorded to examine whether pitch accent affects consonant perturbation on the first syllable in disyllables. These first syllables were all CV, where C=[t, d], and V=[a, e, o] and were all followed by the same consonant [k]. In total, the stimuli consisted of 702 monosyllables (18 monosyllables * 3 repetitions * 13 speakers) and 468 disyllables (6 disyllables * 2 pitch accent patterns * 3 repetitions * 13 speakers) (Table A2 and A3. The speakers were instructed in Japanese and English, and were recorded in a quiet room using Audacity on a laptop connected with an Azden ECZ-990 microphone, with a sampling rate of 44.1 kHz.

2.2. Extraction of f0 and normalization methods

Vowel portions of recorded monosyllables and disyllables were first segmented manually, and f0 values were extracted at 20 normalized time points from each segmented interval using the ProsodyPro Praat script (Xu, 2013).

Segmentation criteria were those described in Jangjamras (2012). For the vowel in each target syllable to be segmented, the vowel onset was defined to be the first zero crossing at the beginning of its voicing in the waveform, and the vowel offset was at the downward zero crossing immediately following the vowel's final glottal pulse (Zhang, Nissen and Francis, 2008).

In order to compare speech production from different speakers, normalization was deemed necessary. In this study, we used the Log Z-score transformation (Laplace, 1820), which is shown by Zhu (1999) to produce a normally distributed range of f0 values.

2.3. Statistical models

2.3.1. Growth curve analysis and functional data analysis of surface f0 contours

Monosyllables and disyllables were separately analysed. For monosyllables, f0 contours on the vowels [a, e, o] after voiceless onsets [p, t, k] are paired with those on the same vowels after voiced onsets [b, d, g]. The same pairing method was applied to the first syllables in disyllables.

We started from a simple model (Mirman, Dixon and Magnuson, 2008):

$$Y_{ij} = (\gamma_{00} + \zeta_{0i}) + (\gamma_{10} + \zeta_{1i}) * Time_{ij} + \varepsilon_{ij}$$

where i is the i^{th} pitch contour and j is the j^{th} time point, γ_{00} is the population average value for the intercept, ζ_{0i} models variation of individual's intercept, γ_{10} is the population average value for the slope, ζ_{1i} models variation of individual's slope and ε_{ij} are the error terms. Orthogonal polynomials in the model were used to ensure that the linear and quadratic terms were not correlated (Mirman, 2014: 52). To optimize the model for data of each pair (e.g. voiceless vs. voiced), we included the quadratic term in the fixed effect, and allowed individuals to vary on the quadratic term only when those terms reached significance according to the likelihood ratio tests. After optimizing the model by including all significant terms, we modelled the pairs as two different contours, and then compared it to a model that treats them as the same, using a likelihood ratio test to determine whether the two models differ significantly. A significant difference between the models indicates a significant difference between the two f0 contours being compared.

In order to examine specific locations of significant perturbation effects, we applied functional data analysis. First, for each utterance produced by each speaker, we fit pairs of normalized f0 curves extracted from the segmented vowel produced after the two types of onset consonants with the following model:

$$y_i(t_j) = f_i(t_j) + \varepsilon_{ij}$$

where $y_i(t_j)$ is the normalized f0 value at time point t_j for the utterance i by each individual and $i=1, \ldots, n$ and $j=1, \ldots, m$. The error term ε_{ij} follows a normal distribution N(0, σ^2). A basis function expansion for $f_i(t_j)$ can be used to fit discrete observations in the form:

$$f_i(t) = \sum_{k=1}^{K} c_{ki}\, \varphi_k(t) = c_i{}'\varphi(t) = \varphi(t)'\, c_i$$

where c_{ki} is the coefficient for the k^{th} basis function used to model the i^{th} utterance, which can be re-written as a K-vector of coefficients of the i^{th} utterance, c_i, and a K-vector of basis function, $\varphi(t)$. The model can be re-written as:

$$Y = \Phi C + \varepsilon$$

where Y is the m×n matrix of observed f0 values for each utterance at each time point, Φ is the m×K matrix of basis functions, C**,** is the K×n matrix of coefficients and $\boldsymbol{\varepsilon}$ the m×n matrix of the error terms. To compute C, we minimized the penalised least squares in the form

$$(Y - \Phi C)'\ (Y - \Phi C) + \lambda C'\left[\int D^2\varphi(s) D^2\varphi'(s)ds\right]C$$

where $D^2\varphi(s)$ is the second derivative of the vector of basis functions $\varphi(t)$. Details of the model can be found in Ramsay and Silverman (2005, 2009) and Frøslie et al. (2013).

For each pair of surface f0 contours (after voiced and voiceless onsets), we chose 20 break points, and used four B-spline basis functions to fit surface f0 curves, and the optimal smoothing values, λ, were determined by the generalized cross-validation measure (GCV). After fitting all f0 curves, we conducted a functional *t*-test to compare the differences between the two f0 curves in each pair. A functional *t*-test constructs a null distribution by randomly shuffling the labels of the two curves. We used 200 random samples, and calculated observed t-statistic, point-wise 0.05 critical value and maximum 0.05 critical value. Statistical significance is reached when observed t-statistics exceed critical values. These two methods were used to model surface f0 contours of both monosyllables and the first syllable of disyllabic words.

2.3.2. Statistical models of underlying pitch targets

In addition to testing consonant onset effects on surface f0 contours, we also tested possible consonantal influence on underlying pitch targets. As mentioned earlier, the underlying pitch target of a functional pitch unit can be modelled quantitatively (Xu and Wang, 2001). Based on this conceptual framework, Prom-On et al. (2009) developed the quantitative target approximation (qTA) model to generate speech f0 contours. The model is shown to successfully generate f0 contours very similar to natural f0 contours. They compare root mean square error (RMSE) and Pearson's correlation coefficient of the original f0 data to those of the synthesized f0, showing that a critically damped linear system is mathematically simpler than an overdamped system with a similar RMSE. Additionally, the critically damped system produces a higher degree of correlation than the overdamped system. In addition, they recommend employing at least a second order linear system, since the vocal fold tension is controlled by two antagonistic muscle forces, and influenced by minor laryngeal muscles, as well as subglottal pressure, which in turn raise and lower f0. However, there is no significant improvement from the third to the fourth order, and the third order is helpful in guaranteeing smoothness across syllable boundaries. Since we are only interested in phonetic and phonologised perturbation effect on a single syllable in this study, the third order was deemed unnecessary.

First, we need to decide the order of the linear system to model the data.

The second-order linear system used by Sun (2001) to estimate underlying pitch targets is shown below:

$$T(t) = at + b$$

$$y(t) = \beta exp^{-\lambda t} + at + b$$

where *T(t)* represents the underlying target, and *y(t)* represents surface f0 values. When $t=0$, the coefficient β is the distance between f0 contour and the underlying pitch target. The parameter λ represents the rate of target approach. Wong (2006) uses a similar model to predict underlying pitch targets for Cantonese tones. Prom-On et al. (2009) chose a third order critically damped system, which constrains the variable control parameters. The model has the form:

$$x(t) = mt + b$$

$$f_0(t) = \left(c_1 + c_2 t + c_3 t^2\right) exp^{-\lambda t} + x(t)$$

where $f_0(t)$ is the response of frequency, and the underlying pitch target is $x(t)$, and λ represents the rate of target approach. The three parameters are determined by initial f0 values, initial velocity and initial acceleration.

In addition to considering the order of the linear system, we also took the polynomial degree of the underlying pitch target into consideration. Xu (2005) proposes that an underlying pitch target might not be linear in some languages. We determined whether the underlying target is linear or not by using the optimal statistical model.

To optimize the order of the linear system and the degree of the underlying pitch targets, we fit four models using non-linear regression.

1) a simple model (sim_1) of the second order linear system with polynomial of the first degree in the underlying targets

$$y(t) = \beta exp^{-\lambda t} + at + b$$

2) a more complex model (com_1) of the third order linear system with polynomial of the first degree in the underlying targets

$$y(t) = \left(c_1 + c_2 \mathrm{t} + c_3 t^2\right) exp^{-\lambda t} + at + b$$

Table 2
The results of growth curve analysis of Japanese.

Pair	P-value
Monosyllables	$\chi^2(3) = 3057.52$ $p < 0.001$ *
L-H $\sigma 1$	$\chi^2(1) = 11.38$ $p < 0.001$*
H-L $\sigma 1$	$\chi^2(1) = 156.81$ $p < 0.001$*

L-H $\sigma 1$: The first syllable in disyllables with a L-H pitch pattern
H-L $\sigma 1$: The first syllable in disyllables with a H-L pitch pattern

3) a simple model (sim_2) of the second order linear system with polynomial of the second degree in the underlying targets

$$y(t) = \beta exp^{-\lambda t} + dt^2 + at + b$$

4) a more complex model (com_2) of the third order linear system with polynomial of the second degree in the underlying targets

$$y(t) = \left(c_1 + c_2 t + c_3 t^2\right) exp^{-\lambda t} + dt^2 + at + b$$

Nonlinear regression needs to be solved iteratively, which is different from the one-step solution of linear regression. Therefore, initial estimates (guesses) need to be made for all parameters, and nonlinear regression procedure will then improve the fit until the improvement is negligible (Motulsky and Ransnas, 1987). Intelligent initial guesses close enough to the solution help the algorithm find the minimizer (Fox and Weisberg, 2010). In order to make intelligent guesses, we plotted each function and changed the parameters so that the shape is similar to the curve connecting the mean f0 values to obtain initial values, then later fitted these models with the obtained initial values (Chen, 2015b). From the four models, we chose the model with the least AIC (Akaike's Information Criterion), a criterion suited for choosing the model with best fit (Kim and Timm, 2006).

After fitting all the optimized models, we examined significant differences in underlying pitch targets due to perturbation by directly testing the parameters of the underlying pitch targets model fitting. This is typically done by performing discriminant analyses or F-tests on extracted parameters of the fitted models (Andruski and Costello, 2004; Xu and Prom-on, 2014). In this study, the goal was to statistically compare each underlying pitch targets pair to see if they significantly differ from each other rather than classifying them, so discriminant analyses or F-tests were not necessary. Moreover, in testing the coefficients, we did not test parameters from each pair of underlying pitch targets uttered by each speaker because we are interested in obtaining results that can be generalized across speakers. Instead, all parameters were obtained from fitting the optimal model to each speaker's data, and a non-parametric Wilcoxon signed-rank test was performed on the coefficients obtained. This non-parametric test was used as an alternative to a paired *t*-test, whose assumptions of normality may not be met, since extracted parameters do not necessarily follow a normal distribution. All statistical analyses were done using R (Core Team 2013). This procedure was used to model underlying pitch targets of both monosyllables and disyllables. As mentioned earlier, for each disyllable, only the underlying pitch target of the first syllable was modelled.

2.4. Results

Results of the growth curve analysis comparing surface f0 contours after voiced vs. voiceless onsets of both monosyllables and first syllables of disyllables with L-H and H-L pitch accents are shown in Table 2. From this table, we see that both monosyllables and disyllables with L-H and H-L pitch patterns showed a significant perturbation effect. Functional data analysis was then applied to detect the location of significant differences.

Fig. 6 plots the fitted surface f0 contours of monosyllables, disyllables with L-H and H-L pitch accent patterns using 20 break points and four B-spline functions, and Fig. 7 plots the graph generated by functional *t*-tests on Japanese monosyllable and disyllable data. The f0 values for monosyllables produced after voiceless and voiceless onsets were significantly different (observed t-statistic exceeding maximum 0.05 critical value from the onset to about 73% of the vowel, and exceeding point-wise 0.05 critical value from the onset to about 76% of the vowel), though the difference is attenuated toward the end of the vowel. However, no significant differences were detected for the first syllable of the disyllables with L-H pitch accent patterns after the two type of voicing onsets. For the H-L pitch pattern, the middle part of the f0 contours showed significant perturbation effect (observed t-statistic exceeding maximum 0.05 critical value in about 9% $\sim$ 64% of the vowel, and exceeding point-wise 0.05 critical value in about 6% $\sim$ 76% of the vowel). These findings suggest that the perturbation effect of preceding consonants is more salient on Japanese monosyllables than on disyllables. We listed the mean fitted values based on each model up to the initial 75% of the surface f0 contours in Table 3. From Fig. 6 and Table 3, our findings are consistent with the literature that vowels after voiced onsets have lower f0 values than those after voiceless onsets.

Table 3
The mean fitted values based on functional data analysis of Japanese.

Pair/Percentage	Monosyllables		H-L $\sigma 1$	
	Voiced	voiceless	voiced	voiceless
15%	−0.29	0.18	0.28	0.56
30%	−0.18	0.28	0.19	0.59
45%	−0.09	0.22	0.21	0.59
60%	−0.07	0.12	0.25	0.58
75%	−0.18	−0.10	0.30	0.57

H-L $\sigma 1$: The first syllable in disyllables with a H-L pitch pattern

Table 4
The estimated Japanese underlying pitch targets from all speakers.

Pair	a Mean (Standard deviation)	b Mean (Standard deviation)
Monosyllables (Voiceless)	−0.08 (0.06)	2.25 (3.94)
Monosyllables (Voiced)	−0.08 (0.23)	7.14 (11.09)
L-H $\sigma 1$ (Voiceless)	0.07 (0.09)	−3.88 (6.19)
L-H $\sigma 1$ (Voiced)	0.14 (0.14)	−9.72 (10.66)
H-L $\sigma 1$ (Voiceless)	−0.02 (0.09)	1.50 (5.08)
H-L $\sigma 1$ (Voiced)	0.006(0.09)	2.02(6.64)

L-H $\sigma 1$: The first syllable in disyllables with a L-H pitch pattern
H-L $\sigma 1$: The first syllable in disyllables with a H-L pitch pattern

Although the growth curve analysis and functional data analysis could capture variation in the surface f0 contours, they fail to distinguish the fact that Japanese consonantal perturbation effect on f0 contours is phonetic rather than phonological.

Therefore, we proceeded to model underlying pitch targets. After fitting the four models for underlying pitch targets, we found that a simple model (sim_1) with polynomial of the first degree in the underlying targets had the lowest AIC value for monosyllables and disyllables with different pitch accent patterns, indicating that sim_1 model was optimal. Therefore, we fitted a sim_1 model to each speaker's data, and the results are shown in Table 4. Then we tested whether the two coefficients (a and b) characterizing the underlying pitch targets of two f0 contours after voiced vs. voiceless onsets were the same. The results of the estimated values of a and b from aggregated data across speakers and Wilcoxon signed-rank tests are shown in Table 4 and Table 5. The coefficients were not significantly different, suggesting that the under-

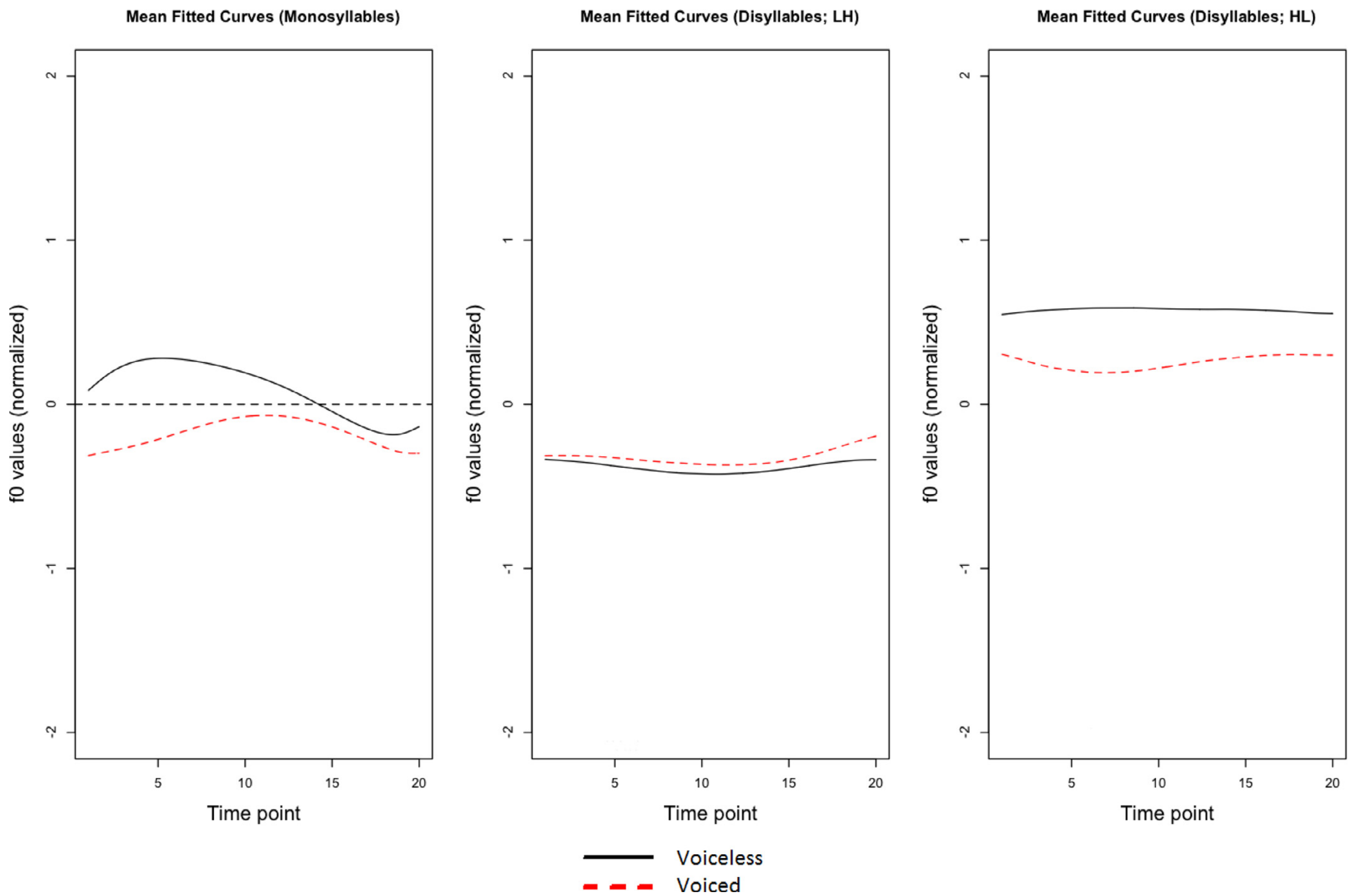

Fig. 6. The mean fitted surface f0 contours of Japanese monosyllables and disyllables based on functional data analysis. The solid black lines stand for mean fitted f0 contours after voiceless onsets, and the dotted red lines stand for mean fitted f0 contours after voiced onsets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
The test results in Japanese underlying pitch targets.

Pair	P-value (parameter a)	P-value (parameter b)	Same or different
Monosyllables	W = 45 p = 1	W = 26 p = 0.19	Same
L-H σ1	W = 20 p = 0.08	W = 73 p = 0.06	Same
H-L σ1	W = 40 p = 0.74	W = 40 p = 0.74	Same

L-H σ1: The first syllable in disyllables with a L-H pitch pattern
H-L σ1: The first syllable in disyllables with a H-L pitch pattern

lying pitch targets are similar after voiced vs. voiceless onsets in both monosyllables and disyllables. In other words, this statistical procedure reveals that phonetic perturbation effects by initial consonants do not significantly affect the underlying pitch targets of Japanese pitch accents.

Fig. 8 plots the mean f0 contours after normalization and the surface f0 contours based on statistical models of underlying pitch targets of both monosyllables and disyllables. From the plot of monosyllables, f0 contours after voiced onsets show some perturbation effect during the first half of the vowel, with f0 convergence toward the end of the vowel. In the plot of disyllables, f0 contours remained relatively constant throughout the vowel, and the contours after voiced vs. voiceless onsets are similar.

In order to eliminate the possibility that a greater degree of variability induced by varied consonants and vowels as well as the carrier phrase in the Japanese data may have led to the observed lack of statistical significance of the underlying pitch targets, we further analysed Japanese target syllables recorded without carrier sentences. We analysed the following CV syllables ([ta] vs. [da]; [te] vs. [de]; [to] vs. [do]) by the same speakers reported in Section 2.1 to make it comparable with Chongming data, which is discussed below. The three pairs of syllables were analysed separately, repeating the non-linear regression fitting procedure and Wilcoxon signed-rank tests explained in Section 2.3.2. Fig. 9 plots the mean f0 contours after normalization of each pair. The fitted mean and standard deviation of estimated underlying pitch targets coefficients for all speakers are shown in Table 6. We proceeded to statistically test those coefficients of pairs with voiced and voiceless onsets, and the results are shown in Table 7. The results are consistent with mixed syllables recorded in carrier sentences, where the underlying pitch targets of each pair did not show significant difference. These results confirm that the underlying pitch targets of pitch accents in Japanese remain unaffected by phonetic perturbation caused by initial consonants both in isolation and in a sentence context.

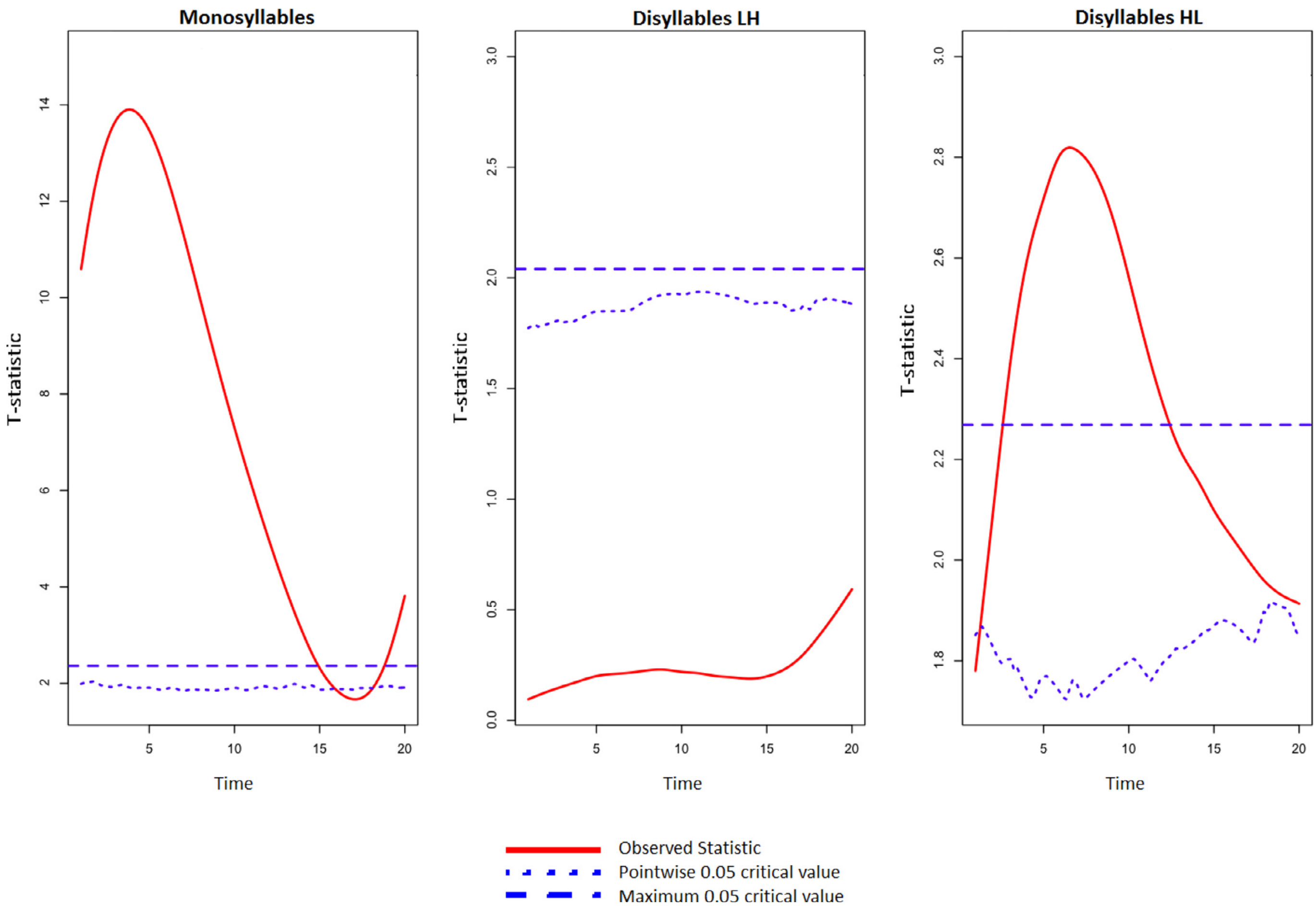

Fig. 7. Functional t-tests of Japanese monosyllables and disyllables. The solid red lines stand for the observed statistics. The dotted blue lines stand for the pointwise 0.05 critical values and the dashed blue lines stand for maximum 0.05 critical values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
The estimated Japanese underlying pitch targets from all speakers.

Pair	a Mean (Standard deviation)	b Mean (Standard deviation)
Monosyllable [ta]	0.07 (0.09)	−4.85 (5.36)
Monosyllable [da]	0.06 (0.17)	−3.56 (10.66)
Monosyllable [te]	0.04 (0.14)	−3.34 (10.17)
Monosyllable [de]	−0.03 (0.17)	1.35 (11.73)
Monosyllable [to]	0.04 (0.1)	−3.01 (5.90)
Monosyllable [do]	−0.02 (0.13)	1.50 (10.03)

L-H σ1: The first syllable in disyllables with a L-H pitch pattern
H-L σ1: The first syllable in disyllables with a H-L pitch pattern

Table 7
The test results in Japanese underlying pitch targets.

Pair	P-value (parameter a)	P-value (parameter b)	Same or different
[ta] vs. [da]	W = 54 $p = 0.59$	W = 40 $p = 0.74$	Same
[te] vs. [de]	W = 33 $p = 0.63$	W = 23 $p = 0.70$	Same
[to] vs. [do]	W = 52 $p = 0.10$	W = 17 $p = 0.17$	Same

L-H σ1: The first syllable in disyllables with a L-H pitch pattern
H-L σ1: The first syllable in disyllables with a H-L pitch pattern

In sum, the perturbation effect in Japanese did not reach significance for underlying pitch targets, indicating that the underlying pitch target remains stable despite phonetic perturbation effect. We applied the same modelling method to Chongming Chinese to test its ability to differentiate putative phonologised perturbation from phonetic, surface perturbation in Chongming Chinese.

3. Acoustic analysis of Chongming Chinese

3.1. Subjects and materials

Thirty native speakers (15 males, 15 females) of Chongming Chinese from 40 to 61 years old were recruited from the city Qidong. All speakers had lived in Qidong for most of their lives, with minimal exposure to other languages and dialects except Mandarin. No participants reported any history of speaking, hearing or language difficulty.

We recorded 1080 monosyllabic word tokens (12 monosyllables * 3 repetitions * 30 speakers). After summarizing data from Zhang (2009), we selected the vowel /æ/ with three onsets /t, t^h, d/ to examine word-initial perturbation. The vowel /æ/ may host all the tones present in the language, removing any potential confound of vowel quality on f0, in particular because vowel height is intrinsically related to f0 (Hombert, 1977; House and Fairbanks, 1953). We used Chinese characters to represent each syllable for elicitation (Table A1).

In order to control for intonational effects, target words are usually embedded in a sentence frame, as in the Japanese study

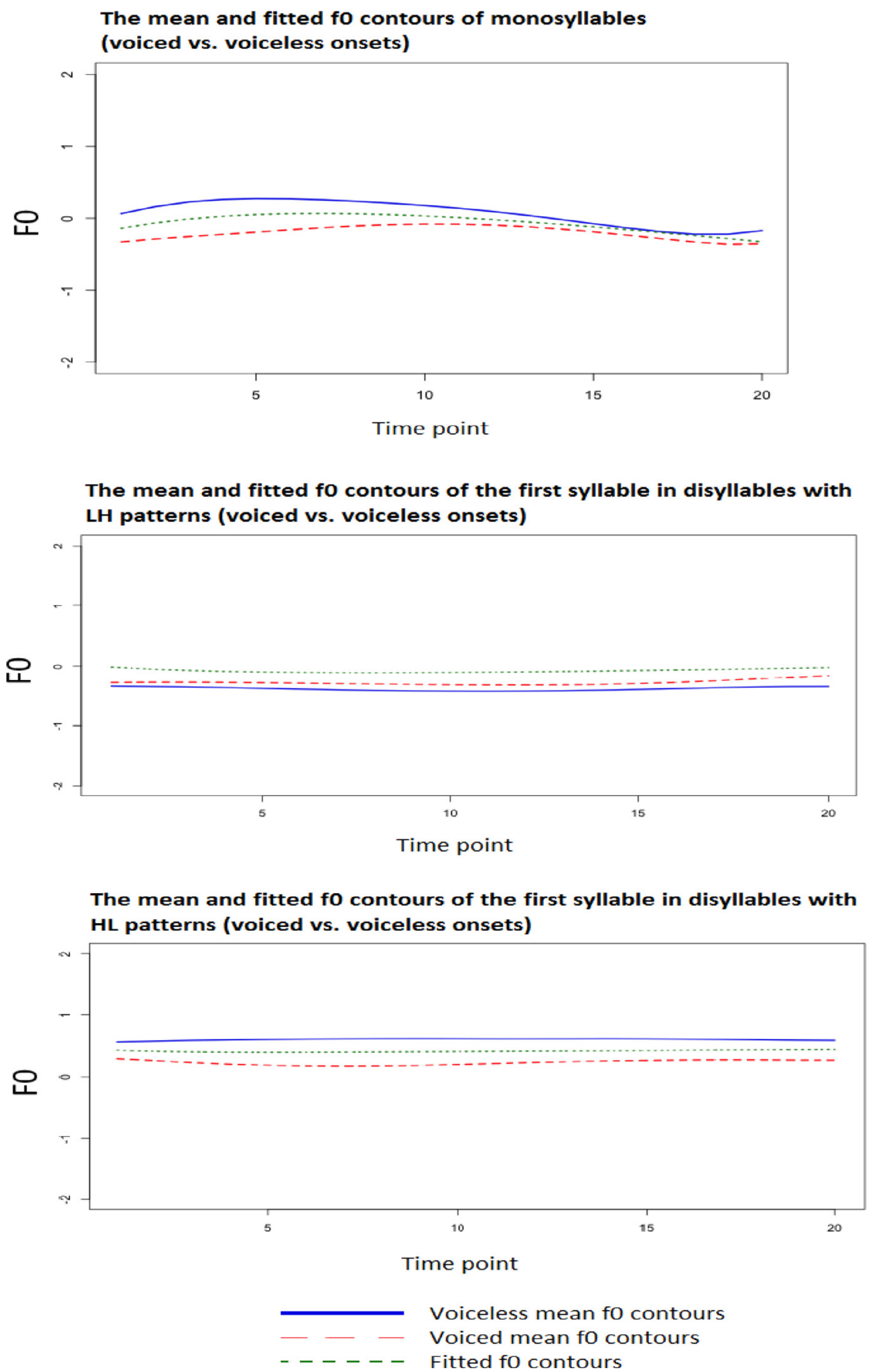

Fig. 8. The fitted and mean f0 contours of Japanese monosyllables and disyllables after voiced vs. voiceless onsets. The solid blue lines stand for mean f0 contours after voiceless onsets. The dotted green lines stand for mean f0 contours after voiced onsets. The dashed red lines stand for the fitted f0 contours. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

above (e.g. Pham, 2003; Sarmah 2009; Xu and Xu, 2003). However, the languages in those studies are either known to have only a few sandhi rules or no reported tone sandhi, so a sentence frame can be designed avoiding sandhi environments. In Chongming Chinese, however, fieldwork shows extensive tone sandhi patterns (Zhang, 2009; Chen and Zhang, 1997), which makes embedding the target word within a larger utterance more challenging. We considered possible carrier sentences for each tone described in Chen and Zhang (1997), which shows that some tones exhibit inevitable sandhi effects regardless of phonetic context. Abramson (1976) claims that the ideal form of a tone is usually considered the isolated monotone, also called the citation tone. We thus recorded words in isolation to remove possible sandhi effects.

The participants were instructed in Chongming Chinese. Recording sessions were conducted in a quiet room using a Marantz PMD 660 digital recorder and a Shure SM2 head-mounted microphone. Recordings were subsequently transferred to a PC with a sampling rate of 48 kHz.

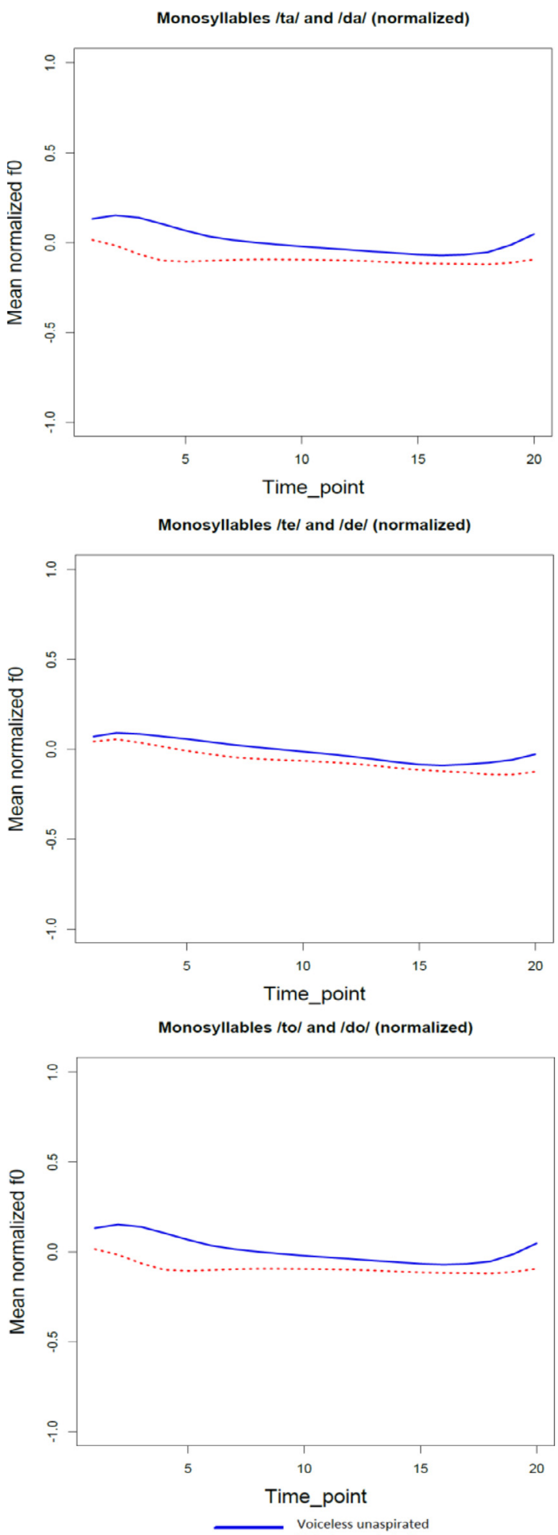

Fig. 9. Mean f0 contours after normalization of each pair: [ta] vs. [da], [te] vs. [de], and [to] vs. [do]. The solid blue lines stand for mean normalized f0 contours after voiceless onsets. The dashed red lines stand for mean normalized f0 contours after voiced onsets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Extraction of f0 and normalization methods

The vowels were first segmented manually, and fundamental frequency (f0) was extracted using a Praat (Boersma and Weenink, 2013) script written by Byunggon Yang, and edited by Jirapat Jangjamras. Time-normalized f0 values were extracted at twenty time points during each vowel, with a 25.6 ms analysis window size. The segmentation criteria and normalization methods were the same as described in Section 2.2.

3.3. Statistical modelling

Before statistical modelling, we also performed acoustic analyses to confirm the existence of the three-way contrast in Chongming Chinese, since it is only reported based on impressionistic data. The results are reported in Appendix B.

As with Japanese, we modelled pairs of surface f0 contours (e.g. after voiceless aspirated (VA) vs. voiceless unaspirated (VU) onsets) using growth curve analysis and examined specific locations of significant perturbation effects by functional data analysis. Also, the underlying pitch targets were investigated by choosing the optimized model. We applied the same statistical techniques in examining both surface f0 contours and underlying pitch targets in Chongming Chinese.

3.4. Results

3.4.1. Statistical modelling of surface f0 contours

Mean f0 contours after normalization of each tonal pair are plotted in Fig. 10. There are some differences in f0 contours after aspirated vs. unaspirated onsets visible in the plot. However, phonologised perturbation effects after voiced vs. voiceless onsets seem to be more dramatic than phonetic perturbation effects after aspirated vs. unaspirated onsets. To determine if surface f0 contours after two different pairs of consonantal onsets: voiceless aspirated vs. voiceless unaspirated (phonetic perturbation), and voiced vs. voiceless unaspirated (phonologised perturbation) are different, we performed growth curve analysis (Mirman, Dixon and Magnuson, 2008; Mirman, 2014: 51–55). The results of the analysis are shown in Table 8.

With the exception of Tone 7 produced after voiceless aspirated and voiceless unaspirated, these results show that f0 contours following both pairs of consonants are significantly different. However, as was the case in Japanese, a significant difference between surface f0 contours does not imply a distinction between the underlying targets. We discuss underlying pitch target modelling of Chongming tones in the next section.

3.4.2. Statistical modelling of underlying pitch targets

Considering the order of the linear system and the degree of the underlying pitch targets, we fitted the four models described in Section 2.3.2, and chose the optimal one.

The best model for T3, T4 and T6 showed an underlying pitch target of a polynomial degree, whereas the rest of the tones exhibited a linear underlying pitch target. In order to test whether there were any statistical differences in underlying pitch targets, we fitted a model for each speaker, removing outliers using the Hampel identifier for the coefficients. Parameter estimates are shown in Table 9.

We then tested the coefficients of the underlying pitch targets within pairs. Since the underlying pitch targets for the pair T5 and T6 are of different degrees, where T6 has a second polynomial degree, and T5 is linear, we did not need to test the T5-T6 pair due to this obvious difference in degrees. We proceeded to test the rest of the pairs with respect to differences in onset voicing and aspiration. The results presented in Tables 10 and 11 show that all pairs after voiced vs. voiceless unaspirated onsets have different

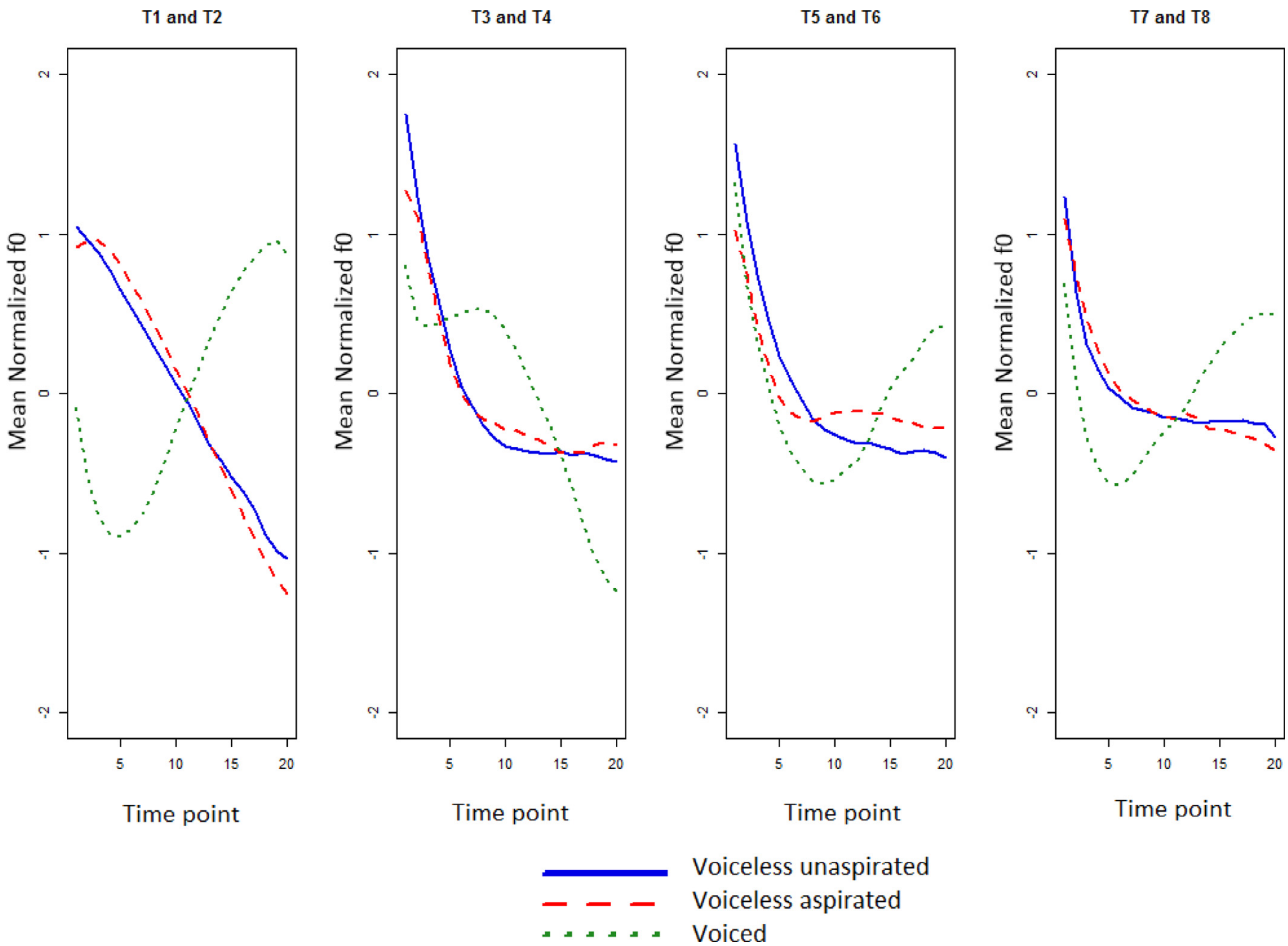

Fig. 10. The plot for mean f0 contours after normalization of each tonal pair: T1 (53) - T2 (24), T3 (435/424) - T4 (241/242), T5 (33) - T6 (213/313) and T7 (55/5) - T8 (23/2). The solid blue lines stand for mean normalized f0 contours after voiceless unaspirated onsets. The dashed red lines stand for mean normalized f0 contours after voiceless aspirated onsets. The dotted green lines stand for mean normalized f0 contours after voiced onsets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 8
The results of growth curve analysis of Chongming Chinese.

Pair	P-value	Pair	P-value
T1/tæ/ vs. T1/tʰæ/	$\chi^2(3) = 31.26$ $p < 0.001^*$	T1/tæ/ vs. T2/dæ/	$\chi^2(3) = 2056.8$ $p < 0.001^*$
T3/tæ/ vs. T3/tʰæ/	$\chi^2(3) = 16.49$ $p < 0.001^*$	T3/tæ/ vs. T4/dæ/	$\chi^2(3) = 456.6$ $p < 0.001^*$
T5/tæ/ vs. T5/tʰæ/	$\chi^2(3) = 63.54$ $p < 0.001^*$	T5/tæ/ vs. T6/dæ/	$\chi^2(3) = 259.57$ $p < 0.001^*$
T7/tæ/ vs. T7/tʰæ/	$\chi^2(3) = 3.94$ $p = 0.27$	T7/tæ/ vs. T8/dæ/	$\chi^2(3) = 250.21$ $p < 0.001^*$

Table 9
The estimated Chongming Chinese underlying pitch targets from all speakers.

Tone	d Mean (Standard deviation)	a Mean (Standard deviation)	b Mean (Standard deviation)
T1/tæ/	NA	−0.08 (0.51)	4.15 (22.79)
T1/tʰæ/	NA	−0.08 (0.46)	9.91 (16.50)
T2/dæ/	NA	−0.77 (1.76)	22.52 (57.44)
T3/tæ/	−0.02 (0.03)	1.26 (1.90)	−31.01 (41.98)
T3/tʰæ/	−0.01 (0.02)	0.90 (1.51)	−20.81 (38.00)
T4/dæ/	−0.001 (0.05)	−0.69 (2.93)	23.25 (73.22)
T5/tæ/	NA	−0.43 (1.10)	14.77 (42.13)
T5/tʰæ/	NA	0.22 (0.98)	−7.39 (36.52)
T6/dæ/	-0.04 (0.06)	2.66 (4.21)	−67.95 (109.89)
T7/tæ/	NA	−0.25 (0.45)	−10.50 (17.24)
T7/tʰæ/	NA	−0.13 (0.48)	−7.53 (18.45)
T8/dæ/	NA	−0.01 (0.52)	−10.99 (21.81)

Table 10
The test results of Chongming Chinese underlying pitch targets with voiceless unaspirated and voiced onsets.

Pair	P-value (differences in parameter d)	P-value (differences in parameter a)	P-value (differences in parameter b)	Same or different underlying pitch targets
T1/tæ/ T2/dæ/	NA	$W=285$ $p=0.29$	$W=403$ $p<0.001^*$	Different
T3/tæ/ T4/dæ/	$W=179$ $p=0.28$	$W=358$ $p=0.009^*$	$W=86$ $p=0.002^*$	Different
T7/tæ/ T8/dæ/	NA	$W=89$ $p=0.002^*$	$W=267$ $p=0.49$	Different

Table 11
The test results of Chongming Chinese underlying pitch targets with voiceless unaspirated and voiceless aspirated onsets.

Pair	P-value (differences in parameter d)	P-value (differences in parameter a)	P-value (differences in parameter b)	Same or different
T1/tæ/ T1/tʰæ/	NA	$W=248$ $p=0.76$	$W=179$ $p=0.28$	Same
T3/tæ/ T3/tʰæ/	$W=200$ $p=0.52$	$W=274$ $p=0.40$	$W=177$ $p=0.26$	Same
T5/tæ/ T5/tʰæ/	NA	$W=92$ $p=0.06$	$W=234$ $p=0.06$	Same
T7/tæ/ T7/tʰæ/	NA	$W=162$ $p=0.15$	$W=237$ $p=0.94$	Same

underlying pitch targets, whereas the underlying pitch targets after voiceless unaspirated vs. voiceless aspirated onsets are the same. These results are consistent with the fieldwork records (Chen and Zhang, 1997), which reported that voiced and voiceless onsets result in differing tone patterns. In short, because the statistical modelling implemented herein accords with previous reports, we conclude that this method successfully modelled phonologised perturbation in Chongming Chinese.

In sum, without optimizing the model to assess underlying pitch targets, statistical modelling of surface f0 contours alone failed to differentiate categorical (phonologised) from gradient (phonetic) perturbation. This suggests the importance of modelling underlying pitch targets for distinguishing between gradient and categorical perturbation effects.

Table 12
Mean fitted values based on functional data analysis of Chongming tones.

Tone/Percentage	5%	10%
T1/tæ/	1.16	1.10
T1/tʰæ/	0.93	0.94
T3/tæ/	1.76	1.57
T3/tʰæ/	1.33	1.28
T5/tæ/	1.57	1.39
T5/tʰæ/	1.05	0.98
T7/tæ/	1.23	1.01
T7/tʰæ/	1.10	0.98

3.4.3. Variations in phonetic perturbation effects across tones

Although growth curve analysis cannot successfully capture differences between phonetic and phonologised perturbation effects, it indicates that the phonetic perturbation may differ across tones. Recall that surface f0 contours after voiceless unaspirated and voiceless aspirated for T7 did not differ significantly. To explore specific regions where phonetic perturbation effect reaches significance, functional analysis was performed using the generalized cross-validation (GCV) measure to determine optimal value of the smoothing parameter λ. An example plot of GCV values against the values of λ when fitting surface f0 contours of T1 after voiceless unaspirated onsets is provided in Fig. 11. The same fitting procedure was applied to T1, T3, T5 and T7. Fig. 12 plots the mean fitted surface f0 contours of these tones (VA vs. VU) together. Functional *t*-tests were then conducted to compare f0 curves after these two types of onsets. Fig. 12 plots the graph generated by functional *t*-tests of each tone. From Fig. 13, T7 did not reach any significance, since the solid red line standing for the observed t-statistic did not exceed the dotted or dashed lines indicating point-wise 0.05 critical values and the maximum 0.05 critical value. For T1, the observed test statistic on the first two calculated points was greater than the permutation critical value for the point-wise 0.05 statistic, but not the maximum 0.05 critical value, showing marginally significant phonetic perturbation on the initial 2% of the curve. T3 and T5 showed significant differences at the beginning of the curves, where the phonetic perturbation on T3 occurred over a small portion of the vowel (the observed t-statistic exceeding point-wise 0.05 critical value for about the initial 2%, and exceeding the maximum 0.05 critical value for about the initial 1%). However, T5 exhibited a larger perturbation effect (the observed t-statistic exceeding point-wise 0.05 critical value for about the initial 5%, and exceeding the maximum 0.05 critical value for about the initial 4%).

Since the perturbation effect did not reach significance after the initial 5% of the f0 contours, we only listed the mean fitted values based on each model at the initial 5% and 10% in Table 12. Consistent with previous studies (Francis et al., 2006; Gandour, 1974; Xu and Xu, 2003), the results of fitted values showed that vowels following voiceless aspirated onsets generally have lower f0 values than those following voiceless unaspirated onsets, though perturbation effect with respect to aspiration is not consistent cross-linguistically.

A summary of results by modelling procedures applied to Japanese and Chongming Chinese is provided in Table 13. Functional data analysis and growth curve analysis produced similar results except for Japanese disyllables with L-H pitch accent, indicating that growth curve analysis may be more sensitive to small differences between curves. However, neither of the two methods testing surface f0 contours is sufficient to differentiate phonetic perturbation from phonologised perturbation, as both models assess only surface tonal patterns. The procedure of testing underlying pitch targets is shown to be both helpful and necessary for

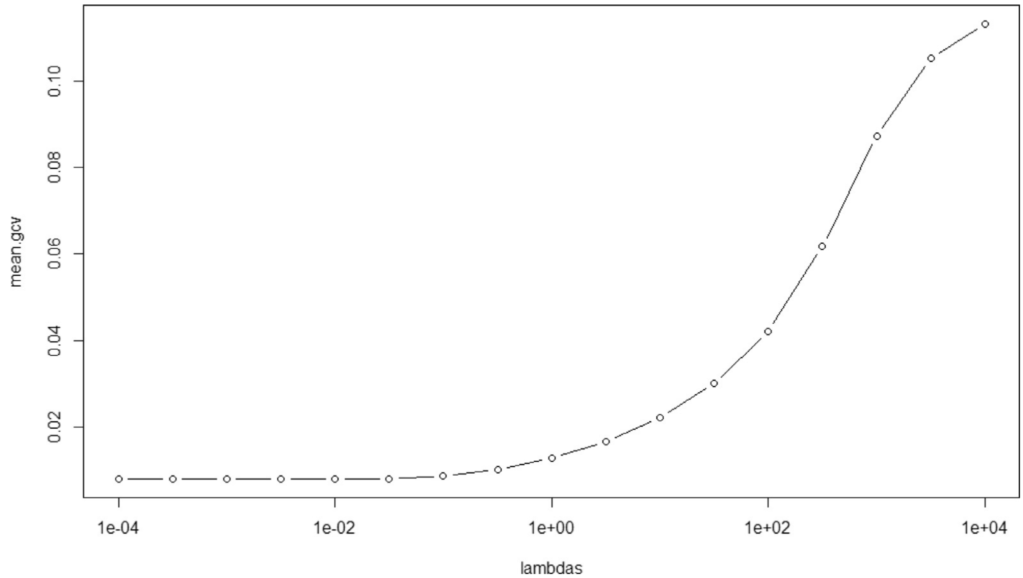

Fig. 11. Generalized cross-validation measures against values of the smoothing parameter.

Table 13
Summary of results by different statistical methods.

Language	Pair	Growth Curve Analysis	Functional Data Analysis	Underlying Pitch Targets
Japanese	Monosyllables (Voiced vs. Voiceless)	Different	Different	Same
	L-H $\sigma 1$ (Voiced vs. Voiceless)	Different	Same	Same
	H-L $\sigma 1$ (Voiced vs. Voiceless)	Different	Different	Same
Chongming Chinese	T1 (VU vs. VA)	Different	Different (marginally significant)	Same
	T3 (VU vs. VA)	Different	Different	Same
	T5 (VU vs. VA)	Different	Different	Same
	T7 (VU vs. VA)	Same	Same	Same
	T1 vs. T2 (VU vs. V)	Different	N.A	Different
	T3 vs. T4 (VU vs. V)	Different	N.A	Different
	T5 vs. T6 (VU vs. V)	Different	N.A	Different
	T7 vs. T8 (VU vs. V)	Different	N.A	Different

VU: Voiceless Unaspirated; VA: Voiceless Aspirated; V: Voiced

determining perturbation effects in Japanese and Chongming Chinese.

4. Discussion

This study examined phonetic and phonologised perturbation effects in Japanese and Chongming Chinese. Section 4.1 summarizes regions where perturbation reached significance, and the relationship between perturbation and pitch accent or tonal types in these two languages. The examination of Chongming Chinese onsets helped us understand acoustic differences among the onsets and the stages of phonologisation. Section 4.2 discusses and compares statistical methods in differentiating phonetic and phonological perturbation and the strength of each method. Section 4.3 discusses future applications of these tested statistical methods as well as limitations of the current study.

4.1. Perturbation in Japanese and Chongming Chinese

We first applied the method to compare the underlying pitch targets of two f0 contours in a non-tonal language, Japanese, and found no significant differences in either monosyllables or disyllables. These results are consistent with the fact that the two observed f0 contours are surface manifestations of the same pitch accent.

We did, however, find a significant phonetic effect of consonant voicing on surface f0 contours of the following vowels in Japanese. This perturbation effect was the most salient on monosyllables. For disyllables with a H-L pitch accent, the effect was observed from 9% - 64% of the first syllable. On the contrary, no significant perturbation was evident for the first syllables of disyllabic words with an L-H accent based on functional data analysis. Kawasaki (1983) states that f0 peaks later for the H-L accent after a voiced stop than after a voiceless stop. However, this pattern is not observed in the L-H accent, where a voiced stop shows a faster increase in pitch rise (due to the initial dip) than a voiceless stop. Our results are consistent with Kawasaki (1983)'s findings in that f0 contours on the H-L accent pattern are significantly more different than those on the L-H accent pattern. Overall, the perturbation effect from onset consonants is more salient in monosyllables than in disyllables (with the L-H or H-L accent). The observed suppression in disyllables may be due to the need to minimize variation in f0 contours to ensure better perception of accent pattern differences.

Moreover, consonant perturbation is suppressed more in words with the L-H accent than the H-L accent. No significant differences in f0 contours were shown for the L-H accent pattern after

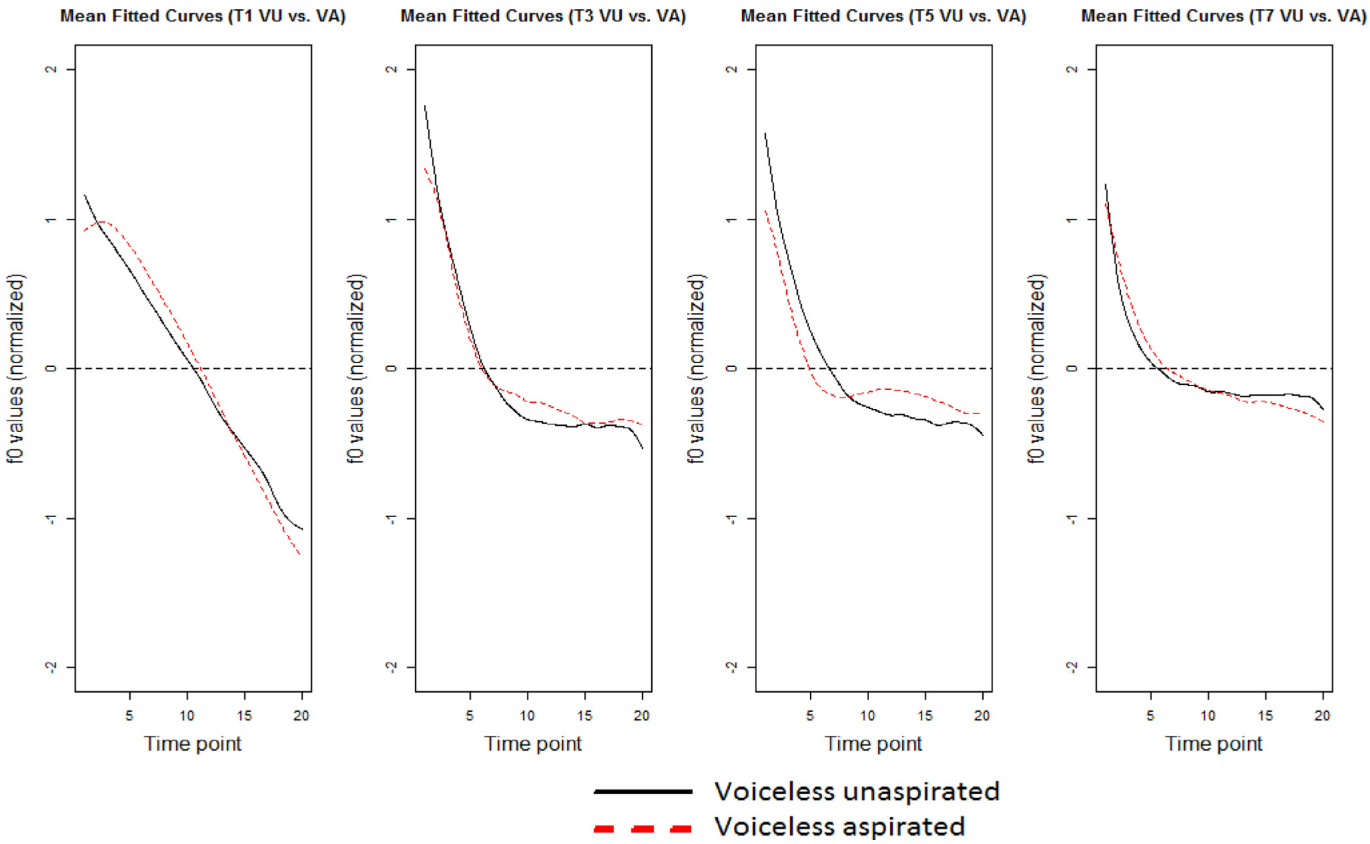

Fig. 12. Mean fitted surface f0 contours of T1, T3, T5 and T7 (VA vs. VU). The solid black lines stand for mean fitted f0 contours after voiceless unaspirated onsets, and the dotted red lines stand for mean fitted f0 contours after voiceless aspirated onsets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

voiced vs. voiceless onsets. However, the suppression did not persist through the entire F0 contours with the H-L accent. The observed significant difference in the middle portion of f0 curves may be due to an interaction with other factors such as contextual variation, also found in many languages (e.g. Mandarin Chinese (Xu, 2005)). Japanese exhibits the phenomenon of pre-low raising (PLR) of surface f0 contours, which is also known as anticipatory dissimilation (Lee, Xu and Prom-on, 2013). Lee et al. (2013) examined various pitch accent patterns and found that a greater peak delay is associated with a lower f0. They propose that the PLR is due to pre-planning of speech, which increases cricothyroid activities in anticipation of a low target. The increase in these activities in turn leads to a higher f0 in an H target as in a H-L pattern. During speech planning of the H-L pattern, which has multiple goals including consonant perturbation suppression, production of an H target, and realization of a higher f0 on the H target due to the PLR effect, the consonant perturbation suppression may be less effective than in the L-H accent patterns.

For Chongming Chinese, underlying pitch targets were found to be different for all tonal pairs with respect to onset voicing, but no significant differences were detected for onsets that differed only by aspiration (i.e. voiceless unaspirated vs. voiceless aspirated). Furthermore, surface f0 contours after onsets differing in aspiration differed across tones, except for T7, and the perturbation effect was more evident during the initial portion of the surface f0 contours. Differences in perturbation effects across tones are also reported in the literature; specifically, low-rising tones are reported to allow greater perturbation than high-rising and high-falling tones in Shanghainese, and rising and low tones show more perturbation than high-rising and high-falling tones in Mandarin Chinese (Chen, 2011; Xu and Xu, 2003). In Chongming, the perturbation effects on the mid tone (T5 (33)) were the most salient. Less notable effects were found on high tones (T1 (53) and T7 (55/5)) or the high-mid-high tone (T3 (435/424)). Mean fitted values showed that f0 values on vowels following voiceless aspirated onsets were generally lower than those following voiceless unaspirated onsets. According to Xu and Xu (2003), subglottal pressure (Ps) varies with the state of vocal folds. To produce higher f0, vocal folds are tenser and may not be affected by aerodynamic factors, leading to smaller perturbation of f0. Though perturbation effect with respect to aspiration is not consistent cross-linguistically (Chen, 2011), our results are consistent with several previous studies (Francis et al., 2006; Gandour, 1974; Xu and Xu, 2003; Zhang, 2009).

4.2. A comparison of results by different statistical methods

In Japanese, surface f0 contours of the same underlying pitch accent showed significant differences, but because perturbation effects of initial consonants are not phonologised, underlying pitch targets were not significantly different. In Chongming, on the other hand, f0 perturbation after voiced vs. voiceless onsets yielded significant differences in both surface contours and underlying pitch targets for all tonal pairs, confirming that, in addition to the surface effects of consonant voicing, the two f0 contours are derived from phonologised f0 representing a tonal contrast. However, for voiceless aspirated vs. voiceless unaspirated onsets, only significant differences in surface f0 contours were observed for three pairs (for the pairs T1/T2, T3/T4, and T5/T6), but the underlying pitch targets were not significantly different, suggesting that perturba-

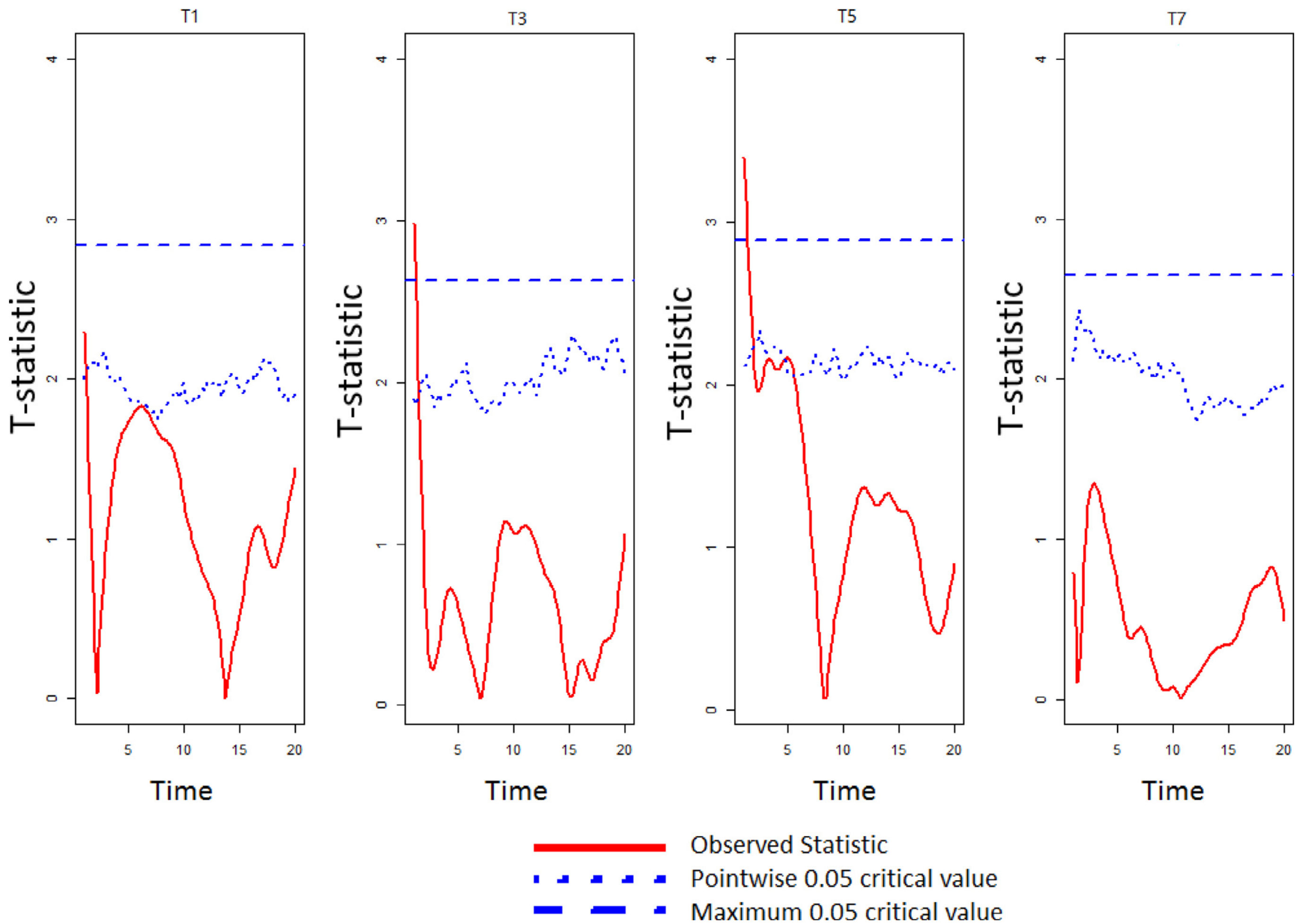

Fig. 13. Functional *t*-tests of T1, T3, T5 and T7. The observed statistics are represented by a solid line, pointwise 0.05 critical values are represented by a dotted line, and the maximum 0.05 critical values are indicated by dashed lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tion effects exerted by these two types of consonants are phonetic in nature.

Our results showed that phonetic and phonologised perturbations are not readily distinguished by examining surface f0 contours alone. By analysing underlying pitch targets in addition to surface contours, though, phonetic and phonologised perturbations are more easily differentiated. As such, statistical modelling of underlying pitch targets is essential for differentiating phonetic perturbation from phonologised f0 changes. The fact that the results were robust for both highly controlled (e.g. in isolation) and variable (e.g. in sentential context) data making it an ideal tool for investigating the underlying pitch targets which tend to be more stable. Xu (2005) uses data from Mandarin Chinese tones to show that the underlying pitch targets are more constant despite variability of surface f0 contours due to contextual variation. Surface f0 curves varied according to the surrounding tones, but they asymptotically converged to a linear target, which is the desired goal for each contrastive tone.

The linear target seems to correspond well to the traditional description of Mandarin tones (high-level, rising and falling) (Chao, 1968), indicating a case where “function and phonology coincide” (Xu, 2005). In the quantitative model of the target approximation conceptual framework (Prom-On et al., 2009), the parameters of target approximation are linked to communicative functions of pitch such as lexical tone, stress and focus, and are the driving force of the linear system. Our modelling showed that phonetic perturbation driven by two different onset types might result in statistically different surface f0 contours, whereas phonologised perturbation (contrastive tones) led to statistical differences in the underlying pitch targets.

In examining surface f0 contours, results by growth curve analysis and functional data analysis are consistent except that surface f0 of Japanese disyllables with the pitch accent L-H, exhibited significant differences in growth curve analysis, but not in functional data analysis. This may suggest that growth curve analysis is more sensitive to small differences in curves. Unlike growth curve analysis, functional data analysis provides us with f0 regions where statistical significance is reached, providing useful information on the time course of phonetic f0 perturbation. We might also be able to use the proportion of regions reaching significance to determine whether perturbation effect is phonologised. However, in order to do so, we have to define an arbitrary proportion of the cut-off point at the boundary between phonetic and phonologised perturbation. By modelling underlying pitch targets, in addition to surface contours, this potential problem is avoided.

4.3. Extending the modelling procedure

As discussed in the introduction, there are two general approaches in dealing with parallel phonetic and phonological phenomena. First is the uni-dimensional approach, which treats phonetics and phonology as one, and accounts for similar phenomena attested in the two subfields only once (Flemming, 2001; Steriade, 2000). The second approach separates phonetics and phonology, recognizing differences in gradient and categorical phenomena (Flemming (2001) also recognizes the difference between the two) attested within and cross languages (Arvaniti, 2007; Chomsky and Halle 1968; Cohn, 2007; Hyman, 2013; Keyser and Stevens 2001; Keating, 1996; Kingston, 2007; Ohala, 1990). Phonologisation of perturbation effects assumes a modular approach, where phonetics and phonology are treated differently (Cohn, 2007). A similar approach has been applied to the investigation of the phenomena of vowel-to-vowel assimilation and vowel harmony. Przezdziecki (2005) examines three dialects of Yoruba, and proposes that vowel-to-vowel coarticulation is phonologised to be vowel harmony, thus supporting Ohala's (1994:491) claim that vowel harmony is a "fossilized remnant" of vowel-to-vowel assimilation. In essence, this suggests that phonetic effects are less substantial phonological effects, or viewed the other way, phonological effects are augmented phonetic effects. Our study also shows that phonetic and phonological processes may exhibit a difference that can be statistically modelled and tested.

From our statistical modelling results, phonologised perturbation showed significant effects on underlying pitch targets for all tonal types, but phonetic perturbation differences did not reach significance for the same underlying targets. The statistical models accurately reflect the current stage of Chongming Chinese where the tonal bifurcation process is complete. Data from languages undergoing phonologisation of f0 will further allow us to test if these modelling methods are effective for detecting the genesis and the time course of tonal phonologisation from phonetic perturbation.

For example, Silva (2006) proposes that Seoul Korean might be developing into a tone language. He argues that VOT has not changed for the tense stops, whereas differences between lax and aspirated stops have decreased. Mean f0 after lax stops is significantly lower than those after tense or aspirated stops, which leads to his conclusion that the contrast between lax and tense stops is manifested as low vs. high tone. A longitudinal case study also confirms these results (Kang and Han, 2013). They showed that in 1935, a Seoul Korean speaker relied almost exclusively on VOT, and 70 years later, the same subject shows more reliance on f0. Perceptually, Seoul listeners also rely on f0 more than VOT to distinguish lenis and aspirated stops, whereas Kyungsang Korean listeners show more reliance on VOT distinction, possibly due to the competitive function of f0 to mark pitch-accent, decreasing the reliability of f0 in differentiating onset consonants (Lee et al., 2013; see also Kirby, 2013). The statistical methods used in the current study have the potential to quantify the change from phonetic perturbation to phonologised perturbation in Seoul Korean.

Moreover, it is reported that tonal and non-tonal languages exhibit differences in the duration of phonetic perturbation effect, with non-tonal languages like American English exhibiting the effect for up to 100 ms (Hombert, 1975). In contrast, this effect is typically shorter in tonal languages, like Thai, with 30 ms for voiceless onsets and 50 ms for voiced onsets (Gandour, 1974), which may be due to a tendency to maximize tonal distinctions by controlling perturbation effects (Hombert, et al., 1979). Our study shows that the proportion of f0 contours of Chongming Chinese affected by phonetic perturbation (i.e. aspiration) is less than that of Japanese. For future studies, it is also possible to quantify and compare perturbation cross-linguistically using statistical modelling procedures, which can reflect the duration where differences in the magnitude and duration of perturbation reach significance.

5. Conclusion

Speech acoustic signals show a great deal of variation, simultaneously conveying their underlying communicative, phonemic functions and their surface contextual phonetic constraints. This study successfully modelled phonetic perturbation of consonant voicing on f0 contours, and phonemic perturbation of contrastive tones, demonstrating that underlying functional units of speech can be extracted from their surface representation; information that may lead to improved algorithms for computerized speech recognition system. The results obtained also yielded further insights into the nature of phonetic and phonologised perturbations as well as their time course, information applicable to future investigation of tonogenesis and other f0 phenomena in the world's languages.

Acknowledgements

We thank Caroline Wiltshire, Andrea Pham, Cynthia Chennault, and participants at the 26th North American Conference on Chinese Linguistics for their comments on an earlier version of the work. We are also indebted to language consultants for their help in this study. Comments and suggestions about statistical modelling and R code from Toby Cheng, Nikolay Bliznyuk and Michael J. Daniels from the Department of Statistics are gratefully acknowledged. We are also indebted to two anonymous reviewers for providing insightful comments and suggestions. This work was supported by Faculty of Humanities at the Hong Kong Polytechnic University [grant number 1-ZVHJ], and partly supported by a grant to the second author from National Natural Science Foundation of China (NSFC: 11504400).

Appendix A. Word list

Table A1
Chongming Chinese monosyllabic word list.

	T1 (55)	T2 (24)	T3 (424)	T4 (242)	T5 (33)	T6 (313)	T7 (ʔ5)	T8 (ʔ2)
C/V	æ							
t	耽		胆		旦		搭	
t^h	毯		坦		探		塔	
d		谈		淡		但		踏
C: consonant; V: vowel; T: tone								

C: consonant; V: vowel; T: tone

Table A2
Japanese monosyllabic word list.

C/V	a	e	o
p	ぱ	ぺ	ぽ
t	た	て	と
k	か	け	こ
b	ば	べ	ぼ
d	だ	で	ど
g	が	げ	ご
C: consonant; V: vowel;			

C: consonant; V: vowel;

Table A3
Japanese disyllabic word list.

Pitch accent: L-H	Pitch accent: H-L
滝 [taki]	凧 [tako]
抱く [daku]	舵機 [daki]
敵 [teki]	梃 [teko]
出来 [deki]	凸 [deko]
徳 [toku]	朱鷺 [toki]
毒 [doku]	土器 [doki]

Appendix B. Phonetic examination of a three-way onset contrast in Chongming Chinese

As mentioned in Section 3.3, we performed acoustic analyses to confirm the existence of the reported three-way contrast of onset consonants before statistical modelling. A better understanding of the acoustic characteristics of onset consonants may also shed light on the phonologisation process. Acoustic cues of consonant voicing as well as phonation types were examined. The data used were the same monosyllables with the vowel /æ/ and three onsets /t, th, d/ as described in Section 3.1.

For consonant voicing, we examined voice onset time (henceforth VOT) (Lisker and Abramson, 1964). For phonation, we used adjusted values of three measurements: H1-H2, H1-A1, and H1-A3. H1 and H2 are the amplitudes of the first and second harmonics of a vowel, and A1 and A3 are the amplitudes of the first and third formants (Gobl and Ní Chasaide, 1992; Gordon and Ladefoged, 2001; Wayland and Jongman, 2003). Hanson (1995) proposed an algorithm of normalizing the measurement of H1-H2. She proposed a correction for the effect of F1 on H1 and H2 by subtracting the amount of

$$20log10\frac{\mathrm{F1}^2}{\mathrm{F1}^2-\mathrm{f}^2}$$

from H1 and H2, where f = f0 for H1 and f = 2f0 for H2. Also the correction of the effect of F1 and F2 on A3 is proposed, where the following amount is added to A3.

$$20\log 10\left(\frac{\left[1-\left(\frac{\mathrm{F3}}{\mathrm{F1}}\right)^2\right]\left[1-\left(\frac{\mathrm{F3}}{\mathrm{F2}}\right)^2\right]}{\left[1-\left(\frac{\mathrm{F3}}{\breve{\mathrm{F}}1}\right)^2\right]\left[1-\left(\frac{\mathrm{F3}}{\breve{\mathrm{F}}2}\right)^2\right]}\right)$$

In this study, we used averaged formants of vowels for the value of $\breve{\mathrm{F}}1$ and $\breve{\mathrm{F}}2$.

The measurements on H1 and H2 were normalized to avoid potential effects from a proximate F1, and similarly A3 was corrected to mitigate a potential boosting effect by the first and second formant (Hanson, 1995). Wayland and Jongman (2003) used normalized *H1 -*H2; *H1 -A1; *H1 -*A3, which has been shown to successfully discriminate between breathy and clear vowels in Khmer. Similar acoustic measurements were also conducted to assess possible phonation types associated with onset consonants. Formants were measured using a Praat script by Christian DiCanio, which extracts mean formant values dynamically across three subintervals within a duration defined by a TextGrid file. The normalized *H1 -*H2; *H1 -A1; *H1 -*A3 were measured using an edited version a Praat script by Bert Remijsen, which makes time-normalized measurements at the onset, middle and offset of the segmented vowel.

VOT was analysed for the reported three-way onset consonants. Voiceless unaspirated stops (VU) have a mean VOT of 12 ms (SD = 6 ms), voiceless aspirated stops (VA) have a mean of 58 ms (SD = 18 ms), and voiced stops (V) have a mean of 15 ms (SD = 9 ms). The 3 ms differences in VOT between voiced and voiceless unaspirated stops may not be quite perceptually meaningful to hearers.

A linear mixed effects model was fitted to the VOT data, with a fixed effect of onset consonants and a random effect of subjects. Subjects were modelled as a random effect. The fixed effect was significant by a likelihood ratio test (χ^2 (1) = 3.98, p = 0.046). Follow-up pair-wise comparisons showed that the VOT was significantly different between all three groups (VU vs. VA: $p < 0.001$; V vs. VU: $p < 0.001$; V vs. VA: $p < 0.001$).

In addition to VOT, three acoustic measurements were analysed: *H1-*H2, *H1-A1 and *H1-*A3. We fitted a mixed effects linear model for each of the three measurements, using position in vowels (onset, mid, and offset), onset consonant type (voiceless aspirated, voiceless unaspirated, and voiced) and an interaction term of these two factors as fixed effects with subjects as a random effect. The fixed effects were tested by likelihood ratio tests, comparing a baseline model without each of these terms. Results showed significant main effects of position in vowels (*H1-*H2: χ^2 (1) = 139.68, $p < 0.001$; *H1-A1: χ^2 (1) = 91.15, $p < 0.001$;

*H1-*A3: χ^2 (1) = 36.77, $p < 0.001$) and onset types (*H1-*H2: χ^2 (1) = 36.91, $p < 0.001$; *H1-A1: χ^2 (1) = 29.28, $p < 0.001$; *H1-*A3: χ^2 (1) = 19.91, $p < 0.001$) in each measurement. The interaction term was also statistically significant for each measurement (*H1-*H2: χ^2 (1) = 5.85, $p = 0.016$; *H1-A1: χ^2 (1) = 36.91, $p < 0.001$; *H1-*A3: χ^2 (1) = 20.70, $p < 0.001$).

A post-hoc analysis was conducted for pair-wise comparisons at each position in vowels. Results showed a significant difference in *H1-*H2 (VU vs. VA: $p < 0.001$; VU vs. V: $p < 0.001$; VA vs. V: $p < 0.001$) and *H1-A1 (VU vs. VA: $p < 0.001$; VU vs. V: $p < 0.001$; VA vs. V: $p < 0.001$) between each pair of onset types only at vowel onset, but not other positions. The measurement *H1-*A3 (VU vs. VA: $p < 0.001$; VU vs. V: $p < 0.001$) showed a pair-wise difference only at the onset of vowels except for voiceless aspirated and voiced onsets (VA vs. V: $p = 0.16$).

In sum, significant differences were present in the three acoustic measurements only at the vowel onset. The measurements suggested diminution of differences toward the vowel offset. These three acoustic measurements successfully distinguished the three types of onset consonants from one another. More interestingly, these results suggest that the state of the glottis differs in the voiced and voiceless unaspirated consonants, which indicates a phonation difference between the two.

In conclusion, all three onset consonant types were distinct in terms of VOT and phonation types in monosyllabic words. The concomitant differences in phonation types among three types of onset consonants at the vowel onset position may directly contribute to the f0 perturbation effects and its phonologisation in a way consistent with the laryngeal model of tonogenesis proposed by Thurgood (2002). The differences may also be due to voice quality retention among "voiced" stops, which is also reported in other languages. DiCanio (2008) summarizes four stages of tonal development based on Kammu data (Svantesson and House, 2006), and also proposes that voice quality of voiced stops is sometimes retained after lowering of tones. In addition, Thompson (1984–5: 40–41) describes that earlier distinctions in voice quality are retained in Vietnamese *ngang* and huyền tones, which lends support to Thurgood (2002)'s laryngeal account.

"Voiced" onset consonants did not show phonetic voicing in isolation (small differences in VOT values compared to voiceless unaspirated stops). A plausible explanation would be that perturbation effect with respect to onset voicing is phonologised, and voiced stops undergo devoicing, as claimed in other studies (DiCanio, 2008; Hyman, 2013; Svantesson and House, 2006).

References

Abramson, A.S., 1976. Thai tones as a reference system. In: Gething, T.W., Harris, J.G., Kullavanijaya, P. (Eds.), Thai Linguistics in Honor of Fang-Kuei Li. Chulalongkorn University Press, Bangkok, pp. 1–12.
Abramson, A.S., Lisker, L., 1985. Relative power of cues: F0 shift versus voice timing. In: Fromkin, V.A. (Ed.), Phonetic Linguistics: Essays in Honor of Peter Ladefoged. Academic, New York, pp. 25–33.
Andruski, J.E., Costello, J., 2004. Using polynomial equations to model pitch contour shape in lexical tones: an example from Green Mong. J. Int. Phonet. Ass. 34, 125–140.
Arvaniti, A., 2007. On the relationship between phonology and phonetics (Or why phonetics is not phonology). In: Proc. ICPhS XVI (Special Session: Between Meaning and Speech: on the Role of Communicative Functions, Representations and Articulations), pp. 19–24.
Bennett, D.M., 1981. Pitch Accent in Japanese: A Metrical Analysis. University of Texas at Austin Publisher.
Blevins, Juliette, 2004. Evolutionary Phonology. Cambridge University Press, Cambridge.
Boersma, P., Weenink, D.. Praat: doing phonetics by computer [Computer program] Version 5.3.51, retrieved 2 June 2013 from.
Chao, Y.R., 1930. A system of tone letters. Le Maitre Phonetique 45, 24–27.
Chao, Y.R., 1968. A Grammar of Spoken Chinese. University of California Press, Berkeley, California.
Chen, M., 2000. Tone Sandhi Patterns Across Chinese Dialects. Cambridge University Press, UK.
Chen, M., Zhang, H.M., 1997. Lexical and postlexical tone sandhi in Chongming. In: Wang, J., Smith, N. (Eds.), Studies in Chinese Phonology. Mouton de Gruyter, Berlin and New York, pp. 13–52.
Chen, S., 2015a. More than F0: an additive model for tonal representations. In: Proc. Int. Conf. on Speech Sci. The Korean Society of Speech Sciences, Seoul.
Chen, S., 2015b. Phonological representations of underlying tonal targets based on statistical modelling in tonal languages. Int. Conf. on Prosodic Studies: Challenges and Prospects (ICPS-1).
Chen, Y.Y., 2011. How does phonology guide phonetics in segment–F0 interaction? J. Phonet 39, 612–625.
Cheng, C.C., Wang, S.-Y., 1977. Tone change in Chaozhou Chinese: a study of lexical diffusion. In: Wang, W.S-Y. (Ed.), The Lexicon in Phonological Change. Mouton, The Hague, pp. 86–100.
Chomsky, N., Halle, M., 1968. The Sound Pattern of English. Harper and Row, New York.
Cohn, A.C., 2007. Phonetics in phonology and phonology in phonetics. Working Papers Cornell Phone. Lab 16, 1–31.
DiCanio, C.T., 2008. The Phonetics and Phonology of San Martín Itunyoso Trique. University of California, Berkeley Ph.D. diss..
Diehl, R.L., Lotto, A.J., Holt, L.L., 2004. Speech perception. Annu. Rev. Psychol. 55, 149–179.
Donnelly, S., 2009. Tone and depression in Phuthi. Lang. Sci. 31, 161–178.
Flemming, E., 2001. Scalar and categorical phenomena in a unifed model of phonetics and phonology. Phonology 18, 7–44.
Fox, J., Weisberg, S., 2010. An R Companion to Applied Regression. Sage.
Francis, A.L., Ciocca, V, Wong, V.K.M., Chan, J.K., 2006. Is fundamental frequency a cue to aspiration in initial stop? J. Acoust. Soc. Am. 120, 2884–2895.
Frazier, M., 2009. Tonal dialects and consonant-pitch interaction in Yucatec Maya. MIT Working Papers Linguist. 59, 59–81.
Frøslie, K.F., Røislien, J., Qvigstad, E., Godang, K., Bollerslev, J., Voldner, N., Henriksen, T., Veierød, M.B., 2013. Shape information from glucose curves: functional data analysis compared with traditional summary measures. BMC Med. Res. Methodol. 13 (1), 1–15.
Gandour, J., 1974. Consonant types and tone in Siamese. J. Phonet. 2, 337–350.
Gibbons, R.D., Hedeker, D., DuToit, S., 2010. Advances in analysis of longitudinal data. Annu. Rev. Clin. Psychol. 6, 79–107.
Gobl, C., Chasaide, A.N., 1992. Acoustic characteristics of voice quality. Speech Commun 11, 481–490.
Gordon, M., Ladefoged, P., 2001. Phonation types: a cross-linguistic overview. J. Phonet. 29, 383–406.
Grabe, E., Kochanski, G., Coleman, J., 2007. Connecting intonation labels to mathematical descriptions of fundamental frequency. Lang. Speech 3 (50), 281–310.
Gubian, M., Francisco, T, Boves, L., 2015. Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. J. Phonet. 49, 16–40.
Hanson, H.M., 1995. Glottal Characteristics of Female Speakers. Harvard University, MA Ph.D. diss..
Haraguchi, S., 1977. The Tone Pattern of Japanese: An Autosegmental Theory of Tonology. Kaitakusha, Tokyo.
Hastie, T., Tibshirani, R., 1986. Generalized additive models. Stat. Sci. (3) 297–318.
Hastie, T., Tibshirani, R., 1990. Generalized Additive Models. Chapman and Hall, London.
Haudricourt, A.G., 1954. De l'origine des tons en vietnamien. J. Asiatique 242, 69–82.
Hildebrandt, K.A. (2003): Manange tones: scenarios of retention and loss in two communities. Ph.D. diss. Department of Linguistics, University of California, Santa Barbara.
Hombert, J.M., 1975. Towards a Theory of Tonogenesis: An Empirical, Physiologically and Perceptually-Based Account of the Development of Tonal Contrasts in Language. University of California, Berkeley PhD. diss..
Hombert, J.M., 1977. Development of tones from vowel height? J. Phonet 5, 9–16.
Hombert, J.M., 1978. Consonant types, vowel, quality and tone. In: Fromkin, V.A. (Ed.), Tone: A Linguistic Survey. Academic Press, New York, pp. 77–111.
Hombert, J.M., Ohala, J.J., Ewan, W.G., 1979. Phonetic explanations for the development of tones. Language 55 (1), 37–58.
House, A.S., Fairbanks, G., 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. J. Acoust. Soc. Am. 25 (1), 105–113.
Howie, J.M., 1974. On the domain of tone in Mandarin. Phonetica 30, 129–148.
Huang, Bufan, 1995. Conditions for tonogenesis and tone split in Tibetan dialects. Linguist. Tibeto-Burman Area 18, 43–62.
Hyman, L.M., 1973a. Consonant types and tone. Southern California Occas. Papers Linguist. No. 1. USC Los Angeles.
Hyman, L.M., 1973b. The role of consonant types in natural tonal assimilations. Southern California Occas. Papers Linguistics 1, 151–179.
Hyman, L.M., 1976. Phonologisation. In: Juilland, A. (Ed.), Linguistic Studies. Anna Libri, Saratoga, Calif, pp. 407–418. presented to Joseph H. Greenberg.
Hyman, L.M., 2013. Enlarging the scope of phonologisation. Origins of Sound Change: Approaches to Phonologisation. Oxford University Press, UK.
Ishihara, S., 1998. Independence between consonantal voicing and vocoid F0 perturbation in English and Japanese. In: Manell, R., Robert-Ribes, J. (Eds.), Proceedings of the 5th International Conference of Spoken Language and Processing, pp. 3107–3110.
Jakobson, R., Halle, M., 1962. Phonology and phonetics. Roman Jakobson Select. Writings 1, 464–504 Mouton.
Jangjamras, J., 2012. Perception and Production of English Lexical Stress by Thai Speakers. University of Florida PhD. diss..

Johnson, K. (2008): Speaker normalization in speech perception. In D.B. Pisoni and R.E. Remez (eds.). The Handbook of Speech Perception, 363–389.
Kang, Y., Han, S., 2013. Tonogenesis in early contemporary Seoul Korean: a longitudinal case study. Lingua 134, 62–74.
Kawasaki, H. (1983): Fundamental frequency perturbation caused by voiced and voiceless stops in Japanese. Cambridge, MA: MIT Research Laboratory of Electronics, Speech Communication Group, Working Papers 3.55–67.
Keating, P.A., 1990. Phonetic representations in a generative grammar. J. Phonetics 18, 321–334.
Keating, P.A. (Ed.), 1994, Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III. Cambridge U. Press.
Keating, P.A., 1996. The phonology-phonetics interface. In: Kleinhenz, U. (Ed.), Interfaces in Phonology. Akademie Verlag, Berlin, pp. 262–278.
Keyser, S.J., Stevens, K.N., 2001. Enhancement revisited. In: Kenstowicz, M., Hale, K. (Eds.), A Life in Language. MIT Press, Cambridge, MA, pp. 271–291.
Kim, K., Timm, N., 2006. Univariate and Multivariate General Linear Models: Theory and Applications with SAS. CRC Press.
Kingston, J., 2007. The phonetics-phonology interface. In: de Lacy, P. (Ed.), The Cambridge Handbook of Phonology, pp. Cambridge University Press, Cambridge, pp. 401–434.
Kirby, J., 2013. The role of probabilistic enhancement in phonologisation. In: Origins of Sound Change: Approaches to Phonologisation. Oxford University Press, Oxford, pp. 228–246.
Kong, E.J. (2009): The development of phonation-type contrasts in plosives: cross-linguistic perspectives. PhD diss. The Ohio State University.
Kubozono, H., 2011. Japanese pitch accent. In: Oostendorp, M.V., Ewen, C., Hume, E., Rice, K. (Eds.). In: The Blackwell Companion to Phonology, 5. Wiley-Blackwell, Malden, MA and Oxford, pp. 2879–2907.
Laplace, P.S., 1820. Théorie Analytique des Probabilités. Mme Ve Courcier, Paris.
Lee, H., Politzer-Ahles, S., Jongman, A., 2013. Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. J. Phonet. 41 (2), 117–132.
Lee, A., Xu, Y., Prom-on, S., 2013. Mora-based pre-low raising in Japanese pitch accent. In: Proc. of Interspeech, Lyon, France, pp. 3532–3536.
Lehiste, I., 1975. The phonetic structure of paragraphs. In: Structure and Process in Speech Perception. Springer, Berlin-Heidelberg, pp. 195–206.
Lehiste, I., Peterson, G.E., 1961. Some basic considerations in the analysis of intonation. J. Acoust. Soc. Am. 33 (4), 419–425.
Li, R., Zhang, H.Y., 1993. Chongming Fangyan Zidain (A dictionary of Chongming Dialect). Jiangsu Education Publishing House.
Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. Psychol. Rev. 74 (6), 431–461. doi:10.1037/h0020279. http://dx.doi.org/.
Liberman, A.M., Mattingly, I.G., 1985. The motor theory of speech perception revised. Cognition 21, 1–36.
Lindblom, B., 1963. Spectrographic study of vowel reduction. J. Acoust. Soc. Am. 35 (11), 1773–1781.
Lisker, L., Abramson, A.S., 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20, 384–422.
Lu, J., 2013. An Investigation of Various Linguistic Changes in Chinese and Naxi. Cambridge Scholars Publishing.
Ma, Y., Mazumdar, M., Memtsoudis, S.G., 2012. Beyond Repeated measures ANOVA: advanced statistical methods for the analysis of longitudinal data in anesthesia research. Region. Anesthesia Pain Med. 37 (1), 99–105.
Matisoff, JamesA., 1973. Tonogenesis in southeast asia. In: Hyman, Larry M. (Ed.), Consonant Types and Tone. University of Southern California, Los Angeles, pp. 71–95.
Mazaudon, M., 2012. Paths to tone in the Tamang branch of Tibeto-Birman (Nepal). In: Gunther, D.V., Seiler, Guido (Eds.), The Dialect Laboratory: Dialects as a Testing Ground for Theories of Language Change. John Benjamins, Amsterdam, pp. 139–178. Studies in Language Companion Series 128.
Mazaudon, M., Michaud, A., 2008. Tonal contrasts and initial consonants: a case study of Tamang, a 'missing link' in tonogenesis. Phonetica 65 (4), 231–256.
McCarthy, J., Prince, A. (1993): Prosodic Morphology I: constraint interaction and satisfaction. MS., University of Massachusetts, Amherst, and Rutgers University, New Brunswick, N. J.
McCawley, J., 1977. Accent in Japanese. In: Hyman, L.M (Ed.), Studies in Stress and Accent [Southern California Occasional Papers in Linguistics 4], pp. 262–272.
Mei, T., 1970. Tones and prosody in middle chinese and the origin of the rising tone. Harvard J. Asia. Stud. 30, 86–110.
Mirman, D., 2014. Growth Curve Analysis and Visualization Using R. Chapman and Hall / CRC.
Mirman, D., Dixon, J.A., Magnuson, J.S., 2008. Statistical and computational models of the visual world paradigm: growth curves and individual differences. J. Mem. Lang. 59 (4), 475–494.
Motulsky, H.J., Ransnas, L.A., 1987. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. FASEB J. 1 (5), 365–374.
Ning, L.H., Shih, C., Loucks, T.M., 2014. Mandarin tone learning in L2 adults: a test of perceptual and sensorimotor contributions. Speech Commun. 63-64, 55–69.
Ohala, J.J., 1990. There is no interface between phonology and phonetics: a personal view. J. Phonet. 18 (2), 153–172.
Ohala, J.J., 1994. Towards a universal, phonetically-based, theory of vowel harmony. In: 3rd Int. Conf. on Spoken Lang. Processing (ICSLP 94), Yokohama, Japan, pp. 491–494. September 18–22.
Öhman, S.E., 1966. Coarticulation in VCV utterances: Spectrographic measurements. J. Acoust. Soc. Am. 39 (1), 151–168.
Pater, Joe, 2009. Weighted constraints in generative linguistics. Cognit. Sci. 33, 999–1035.
Pham, A.H., 2003. Vietnamese tone: a new analysis. Outstanding Studies in Linguistics. Routledge, London.
Prince, A., Smolensky, P., 1993. Optimality Theory: Constraint Interaction in Generative Grammar. John Wiley & Sons.
Prom-On, S., Xu, Y, Thipakorn, B., 2009. Modelling tone and intonation in Mandarin and English as a process of target approximation. J. Acoust. Soc. Am. 125 (1), 405–424.
Przezdziecki, M.A. (2005): Vowel harmony and coarticulation in three dialects of Yoruba: phonetics determining phonology. PhD. diss. Univ. Ithaca.
Pulleyblank, E.G., 1978. The nature of middle chinese tones and their development. J. Chin. Ling. 6, 173–203.
Pulleyblank, E.G., 1991. Lexicon of Reconstructed Pronunciation: In Early Middle Chinese, Late Middle Chinese, and Early Mandarin. UBC press, Vancouver.
Core Team, R, 2013. R: A language and environment for statistical computing. R Found. Stat. Comput.. Vienna, Austria. ISBN 3-900051-07-0 http://www.R-project.org/.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, 2nd Ed Springer-Verlag, New York.
Ramsay, J.O., Silverman, B.W., 2009. Functional Data Analysis with R and MATLAB. Springer-Verlag, New York.
Rose, P., 2002. Independent depressor and register effects in Wu dialect tonology: evidence from Wenzhou tone sandhi. J. Chin. Ling. 30 (1), 39–81.
Sarmah, P. (2009): Tone systems of Dimasa and Rabha: a phonetic and phonological study. PhD. diss. University of Florida.
Shen, Z., 2011. Wu "voiced" stops – a cross-language study. Bull. Chin. Linguist. 5 (2), 49–67.
Shi, B., Zhang, J., 1987. Vowel intrinsic pitch in Standard Chinese. In: Proc. of the 11th International Congress of the Phonetic Science, Tallinn, Estonia, pp. 142–145.
Shih, C., 2001. Generation and normalization of tonal variations. J. Chin. Ling. Monogr. Ser. 17, 32–52.
Shih, C., 2005. Understanding phonology by phonetic implementation. In: Proc. of Interspeech: Lisbon, Portugal, pp. 2469–2472.
Shih, C., Lu, H.Y.D., 2015. Effects of talker-to-listener distance on tone. J. Phonetics 51, 6–35.
Shimizu, K., 1989. A cross-language study of voicing contrasts of stops. Studia Phonologica 23, 1–12.
Shimizu, K., 1994. F0 in phonation types of initial stops. In: Togneri, R. (Ed.), Proc. of the fifth Australian Int. Conf. on Speech Sci. and Tech, pp. 650–655.
Silva, D.J., 2006. Variation in voice onset time for Korean stops: a case for recent sound change. Korean Ling. 13, 1–16.
Steriade, D., 2000. Paradigm uniformity and the phonetics phonology boundary. In: Broe, M., Pierrehumbert, J. (Eds.), Papers in Laboratory Phonology V: Acquisition and the Lexicon. Cambridge University Press, Cambridge, pp. 313–334.
Stevens, K.N., House, A.S., 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. J. Speech Hear. Res. 6 (2), 111–128.
Sun, X.J., 2001. Predicting underlying pitch targets for intonation modelling. 4th ISCA Tutorial and Research Workshop on Speech Synthesis.
Svantesson J., O., House, D., 2006. Tone production, tone perception and Kammu tonogenesis. Phonology 23, 309–333.
Thompson, L.C., 1984–1985. In: O'Harrow, Stephan (Ed.), A Vietnamese Reference Grammar. Mon-Khmer Studies XIII-XIV.
Thurgood, G., 2002. Vietnamese tone: revising the model and the analysis. Diachronica 19, 333–363.
Thurgood, G., 2007. Tonogenesis revisited: Revising the model and the analysis. In: Studies in Tai and Southeast Asian Linguistics. Ek Phim Thai Co., Bangkok, pp. 263–291.
Ting, P.H. (1996): Tonal evolution and tonal reconstruction in Chinese, in Huang, Li, new horizons in Chinese linguistics 141–159. Kluwer Academic Publishers, Dordrecht.
Ullah, S., Finch, C.F., 2013. Applications of functional data analysis: a systematic review. BMC Med. Res. Methodol. 13 (1), 1–43.
Wang, W.S-Y., 1967. Phonological features of tone. Int. J. Am. Ling 33 (2), 93–105.
Wayland, R., Jongman, A., 2003. Acoustic correlates of breathy and clear vowels: the case of Khmer. J. Phonet. 31, 181–201.
Whalen, D.H., Levitt, A.G., 1995. The universality of intrinsic F0 of vowels. J. Phonet. 23, 349–366.
Wieling, M., Fabian, T., Arnold, D., Tiede, M., Harald Baayen, R., 2014. Large-scale analysis of articulatory trajectories using generalized additive modelling. Poster presented at the 10th International Seminar on Speech Production, Cologne.
Wong, Y.W., 2006. Contextual tonal variations and pitch targets in Cantonese. In: Proc. Speech Prosody. Dresden, Germany.
Wood, S., 2006. Generalized Additive Models: An Introduction. R. Taylor and Francis Group, LLC with.
Xu, C.X., Xu, Y., 2003. Effects of consonant aspiration on Mandarin tones. J. Int. Phonet. Ass. 33, 165–181.
Xu, D., Fu, J., 2015. Space and Quantification in Languages of China. Springer International Publishing, Switzerland.
Xu, Y., 2005. Speech melody as articulatorily implemented communicative functions. Speech Commun. 46, 220–251.
Xu, Y. (2013): ProsodyPro —A tool for large-scale systematic prosody analysis; tools and resources for the analysis of speech prosody (TRASP 2013), pp. 7–10. Aix-en- Provence, France. Retrieved from http://www.homepages.ucl.ac.uk/∼uclyyix/ProsodyPro/

Xu, Y., Lee, A., Prom-on, S., Liu, F., 2015. Explaining the PENTA model: a reply to Arvaniti and Ladd. Phonology 32, 505–535.
Xu, Y., Prom-on, S., 2014. Toward invariant functional representations of variable surface fundamental frequency contours: synthesizing speech melody via model-based stochastic learning. Speech Commun. 57, 181–208.
Xu, Y., Wang, Q.E., 2001. Pitch targets and their realization: evidence from Mandarin Chinese. Speech Commun. 33, 319–337.
Ye, X.L., 1983. Wujiang fangyan shengdiao zai diao cha [Wujiang tones revisited]. Fangyan 32–35.
Yip, M., 1980. The Tonal Phonology of Chinese. Garland, New York MIT PhD Dissertation. Published 1990.
Zhang, C., 2009a. Why would aspiration lower the pitch of the following vowel? Observations from Leng-shui-jiang Chinese. The 10th Annual Conference of the International Speech Communication Association (Interspeech 2009), Brighton, U.K.
Zhang, C., Chen, S., 2016. Toward an integrative model of talker normalization. J. Exp. Psychol. 42 (8), 1252–1268.
Zhang, H.Y., 2009b. Chongming Fangyan Yanjiu (The Study of Chongming Chinese). China Social Sciences Press, Beijing.
Zhang, Y., Nissen, S., Francis, A., 2008. Acoustic characteristics of English lexical stress produced by native Mandarin speakers. J. Acoust. Soc. Am. 123 (6), 4498–4513.
Zhou, D. (1324 Yuan Dynasty): The phonology of Zhongyuan.
Zhu, X., 2012. Multiregisters and four levels: a new tonal model. J. Chin. Linguist. 40 (1), 1–17.
Zhu, X.S., 1999. Shanghai Tonetics. Lincom Europa, Newcastle.