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]

[a] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China; sarah.chen@polyu.edu.hk

[b] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China; caicai.zhang@polyu.edu.hk

[c] Department of Linguistics, University of California, San Diego, CA, U.S.; agmccoll@ucsd.edu

Si Chen

Corresponding author

Department of Chinese and Bilingual Studies,

The Hong Kong Polytechnic University,

Hong Kong SAR, China

Tel: 852-5941-4757

Email: sarah.chen@polyu.edu.hk

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 (i.e., tone distinctions). We investigated perturbation effects in a non-tonal language, Japanese and a tonal language, Chongming Chinese. By using traditional methods modelling surface f0 contours, phonetic and phonologised effects cannot be distinguished as both effects reached significance. We therefore statistically modelled and tested the differences in underlying pitch targets, which successfully distinguished between phonetic and phonologised effects. 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.

Keywords: phonologisation, perturbation, functional data analysis, growth curve analysis, underlying pitch targets

1.0 Introduction

1.1 Approaches to analysing gradient and categorical phenomena

Different arguments have been proposed to define phonetics, phonology and their relationship [Arvaniti, 2007; Chomsky and Halle, 1968; Cohn, 2007; Flemming, 2001; Hyman, 2013; Keating, 1996; Keyser and Stevens, 2001; Kingston, 2007; Ohala, 1990; Steriade, 2000]. An early view, expressed by Chomsky and Halle [1968], is that phonology and phonetics are distinct, where the former deals with discrete and categorical entities (phonological representations), and the latter deals with continuous and gradient phenomena. 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. Hale and Reiss [2000] develop this type of modular view to an extreme, proposing a strictly modular approach, where the distinction between phonology and phonetics is conceptualized as that of form and substance.

In contrast, some work in Optimality Theory [henceforth OT; McCarthy and Prince, 1993; Prince and Smolensky, 1993], attempts to unify the analysis of scalar and categorical phenomena [Flemming, 2001; Steriade, 2000]. These works recognize duplication problems in the phonetic and phonological literature. Specifically, many similar phenomena are conventionalized in both phonetics and phonology, and consequently, accounted for twice, such as assimilation in phonology vs. coarticulation in phonetics. The unified (uni-dimensional) framework requires a detailed phonetic representation, and is argued to be economical in accounting for parallel phenomena [Flemming, 2001].

However, Cohn [2007] points out that this type of uni-dimensional approach ignores a distinction between phonetic (gradient) and phonological (categorical) phenomena attested both within a language and across languages. Cohn [1998:30] notes that a process such as lengthening before voiced consonants can be phonetic in one language but phonological in another. Not only can a process be phonetic and phonological in different languages, parallel gradient and categorical effects are attested within one language [Cohn, 1990, 2007], such as nasalization in Sundanese [Cohn, 1993, 1998], vowel devoicing in Japanese [Tsuchida, 1997] and palatalization in English [Zsiga, 1995]. This type of argument is based on the assumption that phonetics and phonology are distinct, where gradient patterns in phonetics are believed to be the implementation of phonological specifications, showing a mapping relationship between the two modules [Chomsky and Halle, 1968; Cohn, 2007; Keating, 1996; Keyser and Stevens, 2001; Kingston, 2007].

The existence of gradient and categorical effects within a single language is an argument against a unified (uni-dimensional) approach in treating phonetics and phonology as one. For example, in Sundanese [Cohn, 1993, 1998], after a nasal the following vowel shows substantial nasal airflow, and nasal airflow drops sharply during a subsequent stop consonant (i.e. nasal harmony is blocked by stops), which is argued to be a categorical effect. However, nasal airflow drops gradually if the subsequent segment is a glide or liquid, manifesting a gradient effect. Cohn [2007] argues that instrumental phonetic data is required in addition to impressionistic data in order to analyse the gradient and categorical effects. The current study argues statistical modelling can further elucidate differences between gradient and categorical effects from instrumental phonetic data.

1.2 Phonologisation of perturbation effects

The perturbation effects of preceding consonants on f0 are widely noted in many languages, both tonal and non-tonal [Hyman, 1973a, b; Gandour, 1974]. Voiceless consonants are associated with higher f0 and voiced consonants with lower f0, being attested in a variety of languages such as 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 voiced vs. voiceless onset consonants (see Chen [2011] for a summary). For Shanghainese [Chen, 2011], Danish [Hanson, 2009], and the Cheju dialect of Korean [Cho, Jun and Ladefoged, 2002], aspirated stops are correlated with higher f0 than lenis, unaspirated or fortis stops. However, for Beijing

Mandarin [Xu and Xu, 2003], Cantonese [Francis et al., 2006] and Thai [Gandour, 1974], aspirated stops are correlated with lower f0 values.

These perturbation effects are claimed to play a role in the development and phonologisation of lexical tones [Chen, 2000; Hombert et al., 1979; Rose, 2002; Svantesson and House, 2006]. Phonologisation has long been conceptualised as the change of a phonetic property (e.g. intrinsic f0 perturbation after voiced/voiceless obstruents) into a phonological one, such as lexical tonogenesis fromengthening before voiced consonants, palatalization, tonogenesis from phonation, and tonal bifurcation from onset voicing changes [Hyman, 2013: 4]. Phonologisation occurs when a phonetic process becomes phonological [Hyman, 1975: 171], and when an intrinsic by-product becomes extrinsic and unpredictable [Hyman, 1976: 408].

Some early studies reported phonologisation of phonetic f0 lowering effects due to voiced obstruents [Hombert, 1978; Hyman and Schuh, 1974]. Diachronically, languages appear to undergo different stages of tonal development [Svantesson and House, 2006]. Svantesson and House [2006] claim that Northern and Western dialects of Kammu (a Mon-Khmer language) make use of f0 to contrast word meaning, but the Eastern dialect does not use f0 but rather voicing contrasts. In essence, Northern and Western dialects of Kammu may be shifting the locus of contrast from onset voicing to tone, as in Yoruba. DiCanio [2008] summarizes four stages of tonal development based on Kammu data: one, the breathiness of voiced stops leading to pitch lowering; two, voiced stops undergoing devoicing; three, voice quality retained as aspiration on consonants in some dialects, and four, the loss of voice quality in some dialects.

The belief that phonologisation of tone derives from obstruent voicing dates back to early studies. Haudricourt [1954] argues that Vietnamese tones developed from initial and final consonants, and Thurgood [2002] updates this segmentally-based account with a laryngeally-based analysis, which is further attested in a variety of languages such as San Martín Itunyoso Trique [DiCanio 2008], Manange [Hildebrandt, 2003], Tamang [Mazaudon and Michaud, 2008, 2012], Middle Chinese [Pulleyblank, 1978], Wujiang Chinese [Ye, 1983]. Thurgood [2002] contends that laryngeal gestures connected with voice qualities rather than a consonant-based account [i.e. Haudricourt, 1965], provide a better avenue for understanding tonogenesis. Thus, voiced onsets may result in breathiness, leading to lowering of pitch due to laryngeal gestures.

Tonal bifurcation from voicing is also reported in Chinese Wu dialects such as Chongming and Songjiang Chinese, where onset voicing conditions tonal split into high and low registers, yielding eight tones in total [Chen and Zhang, 1997; Chen, 2000]. Chen [2000] argues that tone bifurcation is sensitive to voicing at onset, but that tones may additionally split according to aspiration. Bifurcation from aspiration differences is described in Wujiang Chinese [Ye, 1983] and Manange [Hildebrandt, 2003:15], where a tone split is observed with respect to aspiration in obstruent onsets. Accordingly, we investigate the role of onset voicing and aspiration in phonetic and phonologised f0 perturbation in Japanese and Chongming Chinese.

1.3 Quantitative models of surface f0 contours and underlying pitch targets

We employed the conceptual framework of underlying pitch targets [Xu and Wang, 2001] to facilitate investigation of phonetic and phonologised perturbation effects both on surface f0 contours and at a more abstract level, on underlying pitch targets. Xu and Wang [2001] distinguish between underlying pitch targets and surface tonal contours, where articulatory constraints affect the realizations, i.e. surface contours, of a given underlying pitch target. This conceptualization of tonal contours and surface realizations falls out from the distinction between phonology and phonetics suggested above. Within this framework, it is possible to test whether phonetic realizations (i.e. surface contours) as well as phonological representations (i.e. pitch targets) are affected by onset consonants using the statistical modelling methods described below.

Figure 1, 2 and 3 show three hypothesized levels of perturbation effects. In Figure 1, the perturbation effect does not reach statistical significance when the whole surface f0 contour is modelled. In Figure 2, the perturbation effect reaches statistical significance for surface f0 contours after different types of onsets, but the effect on underlying pitch targets does not reach significance. In Figure 3, both surface contours and underlying pitch targets differ significantly. Therefore, we may be able to differentiate phonetic and phonologised perturbation effect either by statistically testing surface f0 contours or underlying pitch targets. Note that Figures 1-3 only represent a conceptual framework, and both surface f0 contours and underlying pitch targets need to be modelled quantitatively to test for statistical significance. 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. Similarly, the underlying pitch targets cannot be inferred easily without optimization of model fitting, and statistical tests need to be conducted for comparisons of these targets.

1.3.1 Surface f0 contours

For testing 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) compares values over a bundle of time points. After fitting the model, the role each predictor plays in the model can be examined separately [Chen, 2015a; Wood, 2006]. Hastie and Tibshirani [1990] further argue for the advantages of additive models that they do not assume linear dependence in predictors as the standard parametric multiple linear regression model does, which may not exist in the data. Since GAM applies to more than one independent variable, and the data in this study only has time point as the independent variable, we did not use this method.

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 in terms of the 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 because it does not model variation of individual’s trajectories over time.

[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.

Functional data analysis is another technique to compare a bundle of time series and captures the correlation between them. 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 using continuous smooth dynamics for accurate parameter estimation since it creates functional data out of discrete observations over time and significant noise reduction due to smoothing [Ullah and Finch, 2013].

In this study, growth curve analysis and functional data analysis were used to model surface f0 contours. Specifically, we used growth curve analysis to investigate whether a perturbation effect may result in statistically significant differences of surface f0 contours, and whether this analysis may differentiate phonetic and phonologised perturbation effect. In addition, functional data analysis was used in this study to obtain specific locations where two curves showed statistical differences [Ramsay and Silverman, 2005; 2009], which provided 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 articulatorally operable units associated with linguistically functional pitch units such as tone and pitch accent”. A pitch target may be a static one, such as [high] or [low], or a dynamic one, such as [rise] or [fall]. In this way, surface f0 contours are viewed as the realization of the underlying pitch targets, or alternatively, as part of the surface acoustics, generated through target approximation (TA) [Xu, Lee, Prom-on, Liu, 2015]. Xu and Wang [2001] incorporate variations in surface f0 contours such as vowel intrinsic f0, perturbation from initial consonants and contextual variations into this conceptual framework. F0 properties were characterized along two dimensions, by functionally specified properties and by articulatorally obligatory properties [Xu, 2005].

In the quantitative target approximation model [Prom-On, Xu and Thipakorn, 2009], these two parts are also modelled separately, where the articulatory mechanism is sequential target approximation and communicative functions are modelled as the driving force of a linear system. In this study, we do not use this framework only for the purposes of synthesizing f0 contours, but we are more interested in statistically testing the differences between underlying pitch targets, which offers the possibility to differentiate phonetic and phonologised perturbation quantitatively. Detailed statistical testing procedures for evaluating underlying pitch targets are provided in Section 2.4.

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, voicing contrast retains in Chongming Chinese as reported in the literature, which is relatively rare in other Chinese dialects. Japanese is an easier case to analyse with only a two-way contrast in the onsets: voiceless vs. voiced, whereas Chongming Chinese has a contrast in aspiration in addition to voicing contrast. Therefore, we first examined perturbation effects in Japanese, and investigated models that may reflect the fact that the phonetic perturbation effect is not phonologised in Japanese. Then the same procedures were applied to Chongming Chinese to differentiate phonetic and phonologised perturbation.

Japanese is a pitch-accent language, which bears a tone on each mora that can be associated by rules once the accent marker is known, and 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. Some early studies have examined perturbation effects in Japanese with 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 60ms after vowel onset following voiced stops. Kawasaki [1983] identifies a 40ms 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 remains consistent. Using a mixed effects logistic regression model, Kong [2009] also shows that f0 values on a Japanese female speaker's tokens contribute significantly in discriminating voiced and voiceless stops. Figure 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, but also 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, 2009], it is suggested that the tone system and tone sandhi behaviour of Chongming Chinese may differ between younger and older speakers [Zhang 2009]. To control for intergenerational differences, the current study focuses on speech productions of older speakers.

Chongming Chinese has eight contrastive tones [Zhang, 2009]. Zhang uses Chao's [1930] 5-point scale to describe the language, where 5 represents the highest tone value, and 1 represents the lowest. The eight tones are described in Table 1, which organizes tones according to middle Chinese tonal categories [Chen and Zhang, 1997]. There were four tonal categories in Middle Chinese: ping "level", shang "rising", qu "departing" and ru "entering", recorded in the dictionary *Qieyun* (AD601) [Chen, 2000], and significantly, Middle Chinese exhibiting voicing contrasts lost in most modern Chinese dialects [Pulleyblank, 1991]. These four tones underwent tonal splits conditioned by voicing of Middle Chinese onset obstruents [Cheng and Wang, 1977], which led to a high (yin) and low (yang) register split [Chen, 2000]. Chongming tones are thus divided into four pairs: T1 (53) - T2 (24), T3 (435/424) - T4 (241/242), T5 (33) - T6 (213/313) and T7 (55/5) - T8 (23/2) [Chen and Zhang, 1997].

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	1 H (53)	3 HMH (435/424)	5 M (33)	7 Hʔ (55/5)
	Low register	2 LM (24)	4 LML (241/242)	6 MLM (213/313)	8 Lʔ (23/2)

Chongming Chinese is reported to have a three-way contrast among onset obstruents (voiceless aspirated, voiceless unaspirated and voiced) [Chen and Zhang, 1997]. This study examines acoustic parameters for this three-way contrast, and we are interested in perturbation effect with respect to both voicing and aspiration contrasts. In addition to consonant voicing, some languages show perturbation effects from breathy voiced onsets. Studies on breathy voiced consonants of tonal Tibeto-Burman languages

[Glover 1970; Hombert et al., 1979] as well as Ohala's [1974] research on breathy consonants in Hindi, a non-tonal language, suggest a stronger effect after breathy voiced onsets than after other voiced obstruents. Therefore, we also conducted acoustic measurements related to phonation types in onset consonants.

Fieldwork records show tonal contrasts for 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] yields complementary environments for the four pairs of tones. 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. Figure 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 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 may be based on the whole syllable inventory and is more consistent with the tonal analysis across Chinese dialects. We adopt the generally-agreed view that the four tone pairs underwent tone split, and perturbation effect with respect to voicing contrasts is exaggerated and phonologised, whereas perturbation with respect to aspiration contrasts is not phonologised. Also, it is possible that subsequent to Chongming Chinese tonal phonologisation surface tonal contours having changed considerably.

1.5 The current study

This study focuses on phonetic and phonologised perturbation effects on f0 following different types of onset consonants in Japanese and Chongming Chinese. Japanese has a two-way contrast: voiced vs. voiceless onsets [e.g. Shimizu, 1989]. Chongming Chinese has a three-way contrast for obstruents: voiced, voiceless aspirated and voiceless unaspirated consonants [e.g. Chen and Zhang, 1997].

We assume that phonologisation of tone results in greater differentiation of f0 than mere phonetic perturbation. More concretely, we assume that these differences in f0 are manifested by changes in surface contours as well as underlying pitch targets. Using Hyman's [1976] proposal that phonologised effects are exaggerated phonetic effects (see also Przezdziecki [2005] for this same proposal for vowel harmony), statistical methods were applied to differentiate the difference between phonetic and phonologised (exaggerated) perturbation effects.

Based on the hypothesized situations in Section 1.3, we explored two possible ways to distinguish phonetic perturbation from phonologised perturbation: surface f0 contours and underlying pitch targets. We show herein that while only surface contours are affected in Japanese, leaving underlying pitch targets unaffected, both surface contours and underlying targets are affected by onset obstruents in Chongming Chinese.

2.0 Acoustic analysis of Japanese

Since Japanese is a pitch-accent language [Bennett, 1981; Haraguchi, 1977; McCawley, 1977], there is no phonologisation of perturbation effects as tones. We examined surface f0 contours and the underlying pitch targets after voiced vs. voiceless onsets in Japanese to compare these statistical procedures.

2.1 Subjects and materials

Thirteen native speakers of Japanese (three male, ten female) from age 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 difficulty. None of the participants reported experience learning tonal languages, though they have 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 (LH) and high-low (HL) accent patterns. Each participant read "CV" monosyllables 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 (LH and HL) were also recorded in order to examine whether pitch accent influences perturbation in Japanese. Perturbation effects of the first syllable in disyllables were examined. First syllables were all CV, where C = [t, d], and V = [a, e, o]. All of the first syllables were followed by the same consonant [k]. In total, the stimuli consisted of 702 monosyllables (18 monosyllables * 3 repetitions * 13 subjects) and 468 disyllables (6 disyllables * 2 pitch accent patterns * 3 repetitions * 13 subjects). The subjects were instructed in Japanese and English, and were recorded using Audacity on a laptop connected with an Azden ECZ-990 microphone, with a sampling rate of 44.1kHz in a quiet room.

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 followed the procedures described in Jangjamras [2012]. Vowel onset was defined to be the first zero crossing at the beginning of voicing in the waveform, and the vowel offset was at the downward zero crossing immediately following the final glottal pulse in the waveform [Zhang, Nissen and Francis, 2008 as cited in Jangjamras, 2012].

In order to compare productions from different speakers, normalization was deemed necessary. In this study, we used the Log Z-score transformation, which is shown by Zhu [1999] to produce a normally distributed range of f0 values [see also Fujisaki, 2003; Nolan, 2003 as cited in Prom-On et al., 2009].

Log Z-score equation

$$z_i = \frac{y_i - m_y}{s_y} = \frac{log_{10} x_i - 1/n \sum_{i=1}^{n} log_{10} x_i}{\sqrt{1/(n-1) \sum_{i=1}^{n} (log_{10} x_i - 1/n \sum_{i=1}^{n} log_{10} x_i)^2}}$$

2.3 Statistical models of surface f0 contours: growth curve analysis and functional data analysis

In order to test if a statistically significant difference exists between a pair of f0 contours, 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 mean for the intercept, ζ_{0i} models variation of individual's intercept, γ_{10} is the population mean 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]. Starting from this model, we optimized 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, tested by likelihood ratio tests. After optimizing the model by including all significant terms, we modelled pairs of f0 contours (e.g. after voiceless vs. voiced) as two different contours. Then we compared this model with a model that treats them as the same, using a likelihood ratio test to test whether these two models differ significantly. A significant difference between models indicate a significant difference between the two relevant f0 contours.

In order to examine specific locations of significant perturbation effects, we applied functional data analysis. First, we fit pairs of curves after two types of onset consonants, based on f0 data extracted in segmented vowels for each utterance by each individual. Let $y_i(t_j)$ be the normalized f0 values at time point t_j for the utterance i by each individual, where $i = 1, \dots, n$ and $j = 1, \dots, m$. The f0 curves were then estimated by fitting the following model

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

where the error term $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$. A basis function expansion for $f_i(t_j)$ can be used to fit discrete observations in the form

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

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 $\boldsymbol{c_i}$ and a K-vector of basis function $\boldsymbol{\varphi}(\boldsymbol{t})$. The model can be re-written as

$$\boldsymbol{Y} = \boldsymbol{\Phi C} + \boldsymbol{\varepsilon}$$

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

$$(\boldsymbol{Y} - \boldsymbol{\Phi C})' \, (\boldsymbol{Y} - \boldsymbol{\Phi C}) + \boldsymbol{\lambda C}'[\int D^2\boldsymbol{\varphi}(\boldsymbol{s})D^2\boldsymbol{\varphi}'(\boldsymbol{s})\, ds]\boldsymbol{C}$$

where $D^2\boldsymbol{\varphi}(\boldsymbol{s})$ is the second derivative of the vector of basis functions $\boldsymbol{\varphi}(\boldsymbol{t})$. Details of the model can be found in Ramsay and Silverman [2005; 2009] and Frøslie, Røislien, Qvigstad, Godang, Bollerslev, Voldner, Henriksen, Veierød [2013].

To fit each pair of surface f0 contours after two types of onsets, we chose 20 break points, and used four B-spline basis functions to fit surface f0 curves. The best values for the smoothing parameter, 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 in f0 curves after different types of onsets. 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 monosyllables and disyllables in Japanese as well as monosyllables in Chongming Chinese. For each disyllable, only f0 values extracted from the first syllable were modelled.

2.4 Statistical models of underlying pitch targets

In addition to testing surface f0 contours, we also tested possible changes to underlying pitch targets due to phonetic and phonologised perturbation effects. As mentioned in the introduction, the conceptual framework of underlying pitch target can be modelled quantitatively. Prom-On et al. [2009] quantify the conceptual model proposed by Xu and Wang [2001], assuming that the quantitative target approximation (TA) model should employ 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 and subglottal pressure, which in turn raise and lower f0. They compared root mean square error (RMSE) and Pearson's correlation coefficient between the original f0 data and synthesized f0, showing that a critically damped linear system is mathematically simpler than overdamped system with a similar RMSE. Additionally, the critically damped system produced higher correlations than the overdamped system. With respect to the order, there is no significant improvement from the third to fourth order, and the third order is shown to be helpful in guaranteeing smoothness across syllable boundaries. In this study, we are interested in phonetic and phonologised perturbation effect on a single syllable, so the third order was not necessary. Therefore, we first chose the order of the linear system to model the data.

The second-order linear system used by Sun [2001] to estimate the underlying pitch target is shown below:

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

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

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 the 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

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

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

where f0(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 degree of the underlying pitch target into consideration. Xu [2005] proposes that an underlying pitch target might not be linear in some languages. We examined whether the underlying target is linear or not by choosing the optimal statistical model.

Accordingly, in order to optimize the model in both 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

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

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

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

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 to 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, and then we fit these models with the obtained initial values [Chen, 2015b]. We chose the model with the least AIC (Akaike's Information Criterion), which is a criterion 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. In comparing tonal shapes or underlying pitch targets, parameters of the fitted models are typically extracted, then discriminant analyses or F-tests are employed to classify tones or statistically test the differences between parameters of tones or underlying pitch targets [Andruski and Costello, 2004; Xu and Prom-on, 2014]. This study needs to test whether each pair of underlying pitch targets is statistically different rather than classify and test three or more tones or underlying pitch targets, which therefore does not require discriminant analyses or F-tests. Moreover, in testing the coefficients, we did not test each pair's parameters of underlying pitch targets uttered by the same speaker. Instead, we fit the optimal model to each speaker to extract all parameters, because we are interested in results with generalizability across speakers. We used a non-parametric Wilcoxon signed-rank test to test the coefficients. 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 [R Core Team 2013]. This procedure was used to model underlying pitch targets of monosyllables and disyllables in Japanese as well as monosyllables in Chongming Chinese. For each disyllable, only the underlying pitch target of the first syllable was modelled.

2.5 Results

First, we performed growth curve analysis to examine whether surface f0 contours differ after voiced vs. voiceless onsets. Table 2 showed the results of growth curve analysis, where both monosyllables and disyllables with LH and HL pitch patterns showed significant perturbation effect. Functional data analysis was then applied to detect the location of significant differences.

Table 2 ***The results of growth curve analysis of Japanese***

Pair	P-value
Monosyllables	$\chi^2(3) = 3057.52$ $p < 0.001$ *
LH σ1	$\chi^2(1) = 11.38$ $p < 0.001$*
HL σ1	$\chi^2(1) = 156.81$ $p < 0.001$*

LH σ1: The first syllable in disyllables with a LH pitch pattern
HL σ1: The first syllable in disyllables with a HL pitch pattern

We chose 20 break points, and used four B-spline basis functions to fit surface f0 curves after voiceless and voiced onsets respectively. Figure 6 plots the fitted surface f0 contours of monosyllables, disyllables with LH and HL pitch accent patterns.

Figure 7 plots the graph generated by functional t-tests of Japanese monosyllables and disyllables. The f0 values were significantly different for monosyllables after different types of onsets (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 LH pitch accent patterns after voiced vs. voiceless onsets. For the HL 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% ~ 64% of the vowel, and exceeding point-wise 0.05 critical value in about 6% ~ 76% of the vowel). These findings indicate that the perturbation effect on Japanese monosyllables is more salient than 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 Figure 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		HL σ1	
	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

HL σ1: The first syllable in disyllables with a HL pitch pattern

Although the growth curve analysis and functional data analysis could capture variation on the surface f0 contours, they cannot reflect the fact that Japanese perturbation effect is phonetic and not phonologised as tones. Therefore, we proceeded to model the 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 sim_1 model was optimal. Therefore, we fitted a sim_1 model for each speaker, and the results are shown in Table 4. Then we tested whether the two coefficients (*a* and *b*) characterizing the underlying targets of two f0 contours after voiced vs. voiceless onsets were the same or not. 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 underlying pitch targets are similar after voiced vs. voiceless onsets in Japanese monosyllables and disyllables. Therefore, this statistical procedure suggests that phonetic perturbation effects do not significantly affect underlying pitch targets in Japanese.

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)
LH σ1 (Voiceless)	0.07 (0.09)	-3.88 (6.19)
LH σ1 (Voiced)	0.14 (0.14)	-9.72 (10.66)
HL σ1 (Voiceless)	-0.02 (0.09)	1.50 (5.08)
HL σ1 (Voiced)	0.006(0.09)	2.02(6.64)

LH σ1: The first syllable in disyllables with a LH pitch pattern
HL σ1: The first syllable in disyllables with a HL pitch pattern

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
LH σ1	W = 20 $p = 0.08$	W = 73 $p = 0.06$	Same
HL σ1	W = 40 $p = 0.74$	W = 40 $p = 0.74$	Same

LH σ1: The first syllable in disyllables with a LH pitch pattern
HL σ1: The first syllable in disyllables with a HL pitch pattern

Figure 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 the voiced onsets showed 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 close to each other.

In sum, the perturbation effect in Japanese did not reach significance for underlying pitch targets, indicating that no phonologisation of perturbation effect may be reflected by no significant differences in underlying pitch targets. We applied the same modelling method to Chongming Chinese to test whether this method can also differentiate putative phonologised perturbation in Chongming Chinese.

3.0 Acoustic analysis of Chongming Chinese

3.1 Subjects and materials

Thirty native speakers (15 male, 15 female) 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 1,080 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.

In order to control for intonational effects, target words are usually embedded in a sentence frame, as in the Japanese study 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 48kHz.

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 window for analysis. The segmentation criteria and normalization methods were the same as described in Section 2.2.

3.3 Phonetic examination of a three-way onset contrast in Chongming Chinese

Chongming Chinese is reported to have a three-way contrast among onset obstruents (voiceless aspirated, voiceless unaspirated and voiced). We examined two acoustic properties related to onset voicing, voice onset time (henceforth VOT) [Lisker and Abramson, 1964] and voice bar [Cao and Maddieson, 1992; Liberman et al., 1956; Potter, Kopp and Green, 1947]. VOT measures the time between the release of a stop and the beginning of glottal pulses [Lisker and Abramson, 1964]. A voice bar is "a low resonance around 200-250 Hz" [Fulop, 2011], which can be seen on the spectrum as bands [Cho, 2007].

As for phonation, Hanson [1995] proposes an algorithm to normalize measurements of H1-H2 and a correction of the effect of F1 and F2 on A3. Wayland and Jongman [2003] used the 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.

3.4 Statistical modelling

As with Japanese, we modelled pairs of surface f0 contours (e.g. after voiceless aspirated (VA) vs. voiceless unaspirated (VU)) using growth curve analysis and examined specific locations of significant perturbation effects by functional data analysis. Recall that in Japanese, results showed a phonetic perturbation could be detected by modelling surface f0 contours. However, when underlying pitch targets were examined, the effect did not reach significance. Thus, in the Chongming data, phonologised perturbation effects are predicted to correspond to significant differences in both surface tonal contours as well as underlying pitch targets.

3.5 Results

3.5.1 Onset consonants in monosyllabic data analysis

Voice onset time (VOT) was analysed for the reported three-way onset consonants. A bar plot of the mean and standard deviation of VOT for each type of consonant is shown in Figure 9, where voiceless unaspirated stops (VU) have a mean VOT of 12ms (SD = 6ms), voiceless aspirated stops (VA) have a mean of 58ms (SD = 18ms), and voiced stops (V) have a mean of 15ms (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. 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, and the mean and standard error for the measurements *H1-*H2, *H1-A1 and *H1-*A3 are presented in Figure 10. The column “Onset”, “Middle” and “Offset” represent three points during segmented vowels where acoustic measurements were conducted. 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]. This study used adjusted versions of these three measurements. The measurements on H1 and H2 were normalized to avoid effects from a proximate F1, and similarly A3 was corrected to mitigate a potential boosting effect by the first and second formant, as proposed by Hanson [1995].

We fit a mixed effects linear model for each of the three measurements, using vowel position (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 vowel position (*H1-*H2: $\chi^2(1) = 139.68$, $p < 0.001$; *H1-A1: $\chi^2(1) = 91.15$, $p < 0.001$; *H1-*A3: $\chi^2(1) = 36.77$, $p < 0.001$) and onset types (*H1-*H2: $\chi^2(1) = 36.91$, $p < 0.001$; *H1-A1: $\chi^2(1) = 29.28$, $p < 0.001$; *H1-*A3: $\chi^2(1) = 19.91$, $p < 0.001$) in each measurement. The interaction term was also statistically significant for each measurement (*H1-*H2: $\chi^2(1) = 5.85$, $p = 0.016$; *H1-A1: $\chi^2(1) = 36.91$, $p < 0.001$; *H1-*A3: $\chi^2(1) = 20.70$, $p < 0.001$).

A post-hoc analysis was conducted for pair-wise comparisons at each vowel position. 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. Therefore, diminution of differences toward the vowel offset suggests that the vowel itself does not differ with respect to phonation types. 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.

3.5.2 Statistical modelling of surface f0 contours

Mean f0 contours after normalization of each tonal pair are plotted in Figure 11. There are some differences in the 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. We are interested in whether phonetic perturbation reaches statistical significance in surface f0 contours and whether phonetic and phonologised perturbation can be discriminated by modelling surface f0 contours in Chongming Chinese.

We performed growth curve analysis [Mirman, Dixon and Magnuson, 2008; Mirman, 2014: 51-55] to compare the surface f0 contours after two pairs of onsets: voiceless aspirated vs. voiceless unaspirated onsets and voiced vs. voiceless unaspirated onsets. The results of the analysis are shown in Table 6.

Table 6 ***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/	$\chi^2(3) = 250.21$ $p < 0.001$*

Both phonetic and phonologised perturbation effects reached statistical significance except for phonetic perturbation on T7. Even though surface contours differed for both phonetic and phonologised perturbation effects, this does not necessarily imply that underlying pitch targets also show significant differences. This was the case in Japanese, as surface contours were distinct but underlying pitch targets did not significantly differ.

3.5.3 Statistical modelling of underlying pitch targets

Considering the order of the linear system and the degree of the underlying pitch targets, we fit the four models described in Section 2.4, 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 tones exhibited a linear underlying pitch target. In order to test whether there are any statistical differences in underlying pitch targets, we fit a model for each speaker, removing outliers using the Hampel identifier for the coefficients. Parameter estimates are shown in Table 7.

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 Table 8 and 9 show that all pairs after voiced vs. voiceless unaspirated onsets have different 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.

Table 7 ***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 8 ***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 9 ***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/	$W= 200$	$W= 274$	$W = 177$	Same

T3/$t^hæ$/	p = 0.52	p = 0.40	p = 0.26	
T5/tæ/	NA	W = 92	W = 234	Same
T5/$t^hæ$/		p = 0.06	p = 0.06	
T7/tæ/	NA	W = 162	W = 237	Same
T7/$t^hæ$/		p = 0.15	p = 0.94	

Based on results in this section and the previous section, phonetic perturbation with respect to aspiration and phonologised perturbation effect with respect to voicing cannot be effectively discriminated by modelling surface f0 contours alone, but can be distinguished by optimizing models of underlying pitch targets and statistically testing their coefficients.

3.5.4 Acoustic analysis of phonetic perturbation effects

After successfully distinguishing phonetic and phonologised perturbation effects in Chongming Chinese, this section reports details about the phonetic perturbation effects. Although growth curve analysis cannot successfully capture the differences between phonetic and phonologised perturbation effect, it indicates that the phonetic perturbation may differ across tones, since T7 does not reach significance. Therefore, functional data analysis was applied to explore specific regions where phonetic perturbation effect reaches significance.

The best values for the smoothing parameter, λ, were determined by the generalized cross-validation measure (GCV). 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 Figure 12.

The same fitting procedure was applied to T1, T3, T5 and T7. Figure 13 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. Figure 14 plots the graph generated by functional t-tests of each tone. From Figure 14, 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, 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 10. 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.

Table 10 ***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

A summary of results by modelling procedures applied to Japanese and Chongming Chinese is provided in Table 11. Functional data analysis and growth curve analysis produced similar results except for Japanese disyllables with LH 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 determining perturbation effects in Japanese and Chongming Chinese.

Table 11 ***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
	LH σ1 (Voiced vs. Voiceless)	Different	Same	Same
	HL σ1 (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;

4.0 Discussion

This study examined phonetic and phonologised perturbation effects in Japanese and Chongming Chinese. Section 4.1 summarizes acoustic results concerning the three-way contrast in Chongming Chinese onsets, regions where perturbation effect reaches 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 testing underlying pitch targets to a non-tonal language, Japanese, and found no significant differences in the underlying pitch targets on monosyllables and disyllables. The lack of a significant effect on underlying pitch targets is construed as evidence that the effect of voiced onsets on f0 has not been phonologised in Japanese.

We did, however, find a significant phonetic effect of voicing on surface f0 contours of following vowels in Japanese. The perturbation effect was the most salient on monosyllables. Disyllables with a HL pitch accent had significant perturbation effects on surface f0 contours from 9% - 64% of the first syllable, but disyllables with the pitch accent LH did not show any significant perturbation on the first syllable vowel. The modelling results are consistent with Kawasaki [1983]'s findings that for the HL accent, f0 peaks later. However, the pitch accent pattern did not show much influence on the initial 9% of f0 curves, suggesting that the f0 curves may not be largely affected by consonant perturbation. The observed significance in the middle part of f0 curves may be due to other factors such as contextual variation also found in many languages such as Mandarin Chinese [Xu, 2005].

Before examining perturbation in Chongming Chinese, a series of acoustic measurements were first conducted to examine whether voicing is lost in initial consonants, and whether phonation types differ in onsets. Based on previous work, Chongming Chinese has a three-way contrast among onset obstruents, voiceless aspirated, voiceless unaspirated and voiced [Chen and Zhang, 1997]. Voice onset time (VOT) showed significant differences, though voiceless unaspirated onsets and voiced onsets only differed by 3 ms. Measurements indicative of open quotient and spectral tilt (*H1-*H2, *H1-A1 and *H1-*A3) also confirmed a distinction between the three types of onset consonants word-initially. The effects of consonant type and vowel position as well as their interaction were statistically significant. During post-hoc analysis, the three acoustic measurements *H1-*H2, *H1-A1 and *H1-*A3 showed significant differences at the onset of vowels word-initially. The results thus suggested that the vowel itself did not differ in phonation, but there was a potential phonation difference between "voiced" and voiceless unaspirated consonants.

"Voiced" onset consonants did not show phonetic voicing in isolation (with no voice bar and small differences in VOT values compared to voiceless unaspirated stops). This is similar to Changyinsha Chinese, which has no voice bar for "voiced" consonants in isolation [Cao and Maddieson, 1992]. 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].

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. voicless aspirated). Moreover, there were some differences in phonation types among three types of onset consonants at the vowel onset position, which may also contribute to the perturbation effects and its phonologisation as in the laryngeal model of tonogenesis proposed by Thurgood [2002].

Furthermore, surface tonal contours after onsets that differed in aspiration differed across tones. No statistically significant perturbation effect was found for T7 after voiceless aspirated vs. voiceless unaspirated onsets by functional data analysis and growth curve analysis, although other tones did show some perturbation effects in the initial portion of 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]. The perturbation effects also differed across tones in Chongming Chinese, where the mid tone (T5 (33)) showed the most salient perturbation. Less notable effects were found on other high tones (T1 (53) and T7 (55/5)) or the high-mid-high tone (T3 (435/424)). The mean fitted values showed that f0 on vowels following voiceless aspirated onsets were generally lower than those following voiceless unaspirated onsets. Xu and Xu [2003] argue that 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].

4.2 A comparison of results by different statistical methods

In Japanese, no statistically significant differences in underlying pitch targets were found, since perturbation is not phonologised as tones. Only surface f0 contours showed significant differences. Perturbation after voiced vs. voiceless onsets showed significant differences in surface contours and underlying pitch targets for all tonal pairs in Chongming Chinese, and is taken as evidence for a phonologised tonal contrast. However, for voiceless aspirated vs. voiceless unaspirated onsets, only significant differences in surface f0 contours were detected for three pairs (for the pairs T1/T2, T3/T4, and T5/T6), but underlying pitch targets were not significantly affected.

Our results showed that phonetic and phonologised perturbation are not readily distinguished by surface f0 contours alone. By analysing underlying pitch targets in addition to surface contours, though, phonetic and phonologised perturbation are more easily differentiated. Therefore, we argue that statistical modelling of underlying pitch targets is essential for differentiating phonetic perturbation from phonologised f0 changes. Underlying pitch targets are potentially helpful in this problem because they tend to be more constant. Xu [2005] used Mandarin Chinese tones to show that 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 tone.

The linear target seems to correspond well to the traditional description of Mandarin tones (high-level, rising and falling) [Chao, 1968]. In the quantitative model of this conceptual framework of target approximation [Prom-On et. al., 2009], the parameters of target approximation are linked to communicative functions such as lexical tone, stress and focus, which are the driving force of the linear system. In phonetic perturbation, the driving forces after two different onset types are less likely to show statistically difference, but the implementation of underlying pitch targets may be affected by onset consonants, which may lead to statistical significance on surface f0 contours. However, for phonologised perturbation as tones, the driving forces are more likely to show statistical differences in underlying pitch targets, and in turn the implementation of them, i.e. the surface f0 contours. This is likely a case where "function and phonology coincide" [Chao, 1968].

In examining surface f0 contours, results by growth curve analysis and functional data analysis are consistent except that Japanese disyllables with the pitch accent LH, exhibited significant differences in surface f0 contours using growth curve analysis, but significant differences were not detected using functional data analysis. This may suggest that growth curve analysis is more sensitive to small differences in curves. Functional data analysis, in contrast to growth curve analysis, can provide us with regions where significance is reached, which is useful for understanding the nature 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 a cut-off boundary, which introduces a problem of arbitrariness. For example, for two f0 contours, we may define that the perturbation effect remains to be phonetic if a proportion of less than 10% in the initial part of a vowel is affected. By modelling underlying pitch targets, in addition to surface contours, this potential problem is avoided.

4.3 Extending the modelling procedure

As mentioned in the introduction, there are two general approaches in dealing with parallel phonetic and phonological phenomena. The 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, and it recognizes differences in gradient and categorical phenomena 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.

From our statistical modelling results, phonologised perturbation showed significant effects on the underlying pitch targets for all tonal types, but phonetic perturbation differences did not reach significance for underlying targets. However, both phonetic and phonologised perturbation showed significant differences in surface f0 contours. The statistical models reflect the current stage of Chongming Chinese, though it is important to note that tonal bifurcation may have occurred a long time ago, and the tonal shapes may have changed considerably from then to those realized in the contemporary language. More data from languages undergoing phonologisation is necessary in order to further test whether the initial stage of phonologisation can also be statistically differentiated from phonetic perturbation.

The statistical modelling used in this study can be potentially applicable to other languages, and may shed light on tonogenesis. 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, Politzer-Ahles and Jongman, 2013; see also Kirby, 2013]. The statistical methods used in the current study can help 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 perturbation effect, with non-tonal languages like American English exhibiting the effect for up to 100ms [Hombert, 1975]. In contrast, this effect is typically shorter in tone languages, like Thai, with 30ms for voiceless onsets and 50ms for voiced onsets [Gandour, 1974]. 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.

5.0 Conclusions

This study hypothesized three situations of perturbation effects and tested whether phonetic perturbation and perturbation phonologised as tones in Chongming Chinese and Japanese, could be differentiated by testing surface f0 contours as well as underlying pitch targets.

The results showed that statistical testing on the underlying pitch targets can effectively differentiate phonetic and phonologised perturbation, which also provided more insight into the differences between phonetic and phonological processes.

This work is the first attempt to quantify the difference between phonetic and phonologised perturbation using current statistical methods. Additionally, this study is the first to phonetically examine three types of onset consonants in Chongming Chinese and their relationship with perturbation of f0 during the following vowel. In addition to outlining a modelling procedure applicable to the general analysis of tonogenesis, this paper provides meaningful data from an understudied Wu dialect.

Acknowledgements

We thank Caroline Wiltshire, Ratree Wayland, 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].

Appendix A

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								

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;			

Table A3 Japanese disyllabic word list

Pitch accent: LH	Pitch accent: HL
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]

References

Abramson, A. S. (1976): Thai tones as a reference system; in T. W. Gething, J. G. Harris, P. Kullavanijaya, Thai linguistics in honor of Fang-Kuei Li, pp. 1-12. Chulalongkorn University Press, Bangkok.

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). Proc. ICPhS XVI (Special Session: Between Meaning and Speech: On the Role of Communicative Functions, Representations and Articulations), 19-24.

Bennett, D. M. (1981): Pitch accent in Japanese: a metrical analysis. University of Texas at Austin Publisher.

Boersma, P.; Weenink, D. (2013): Praat: doing phonetics by computer [Computer program]. Version 5.3.51, retrieved 2 June 2013 from http://www.praat.org

Cao, J.; Maddieson, I. (1992): An exploration of phonation types in Wu dialects of Chinese. J. Phonet. 20: 77–92.

Chao, Y. R. (1930): A system of tone letters. Le maitre phonetique 45: 24-27.

Chao, Y. R. (1968): A Grammar of Spoken Chinese. Berkeley, California: University of California Press.

Chen, S. (2015a): More than F0: An Additive Model for Tonal Representations. Proc. the 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, M. (2000): Tone sandhi patterns across Chinese dialects. UK, Cambridge University Press.

Chen, M.; Zhang, H. M. (1997): Lexical and Postlexical Tone Sandhi in Chongming, in J. Wang, N. Smith, Studies in Chinese Phonology, pp.13-52. Mouton de Gruyter, Berlin and New York.

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 W. S-Y. Wang (ed),*The Lexicon in Phonological Change.* The Hague: Mouton pp. 86-100.

Cho, T., Jun, S. A., Ladefoged, P. (2002): Acoustic and aerodynamic correlates of Korean stops and fricatives. J. phonet. 30(2): 193-228.

Cho, W. R. (2007): The voice-bar after closure of coda consonants in the speech of young children. Master’s thesis. Massachusetts Institute of Technology.

Chomsky, N.; Halle, M. (1968): The sound pattern of English. Harper and Row, New York.

Cohn, A. C. (1990): Phonetic and Phonological Rules of Nasalization. UCLA PhD diss. Distributed as UCLA Working Papers in Phonetics 76.

Cohn, A. C. (1993): Nasalization in English: phonology or phonetics. Phonology 10: 43–81.

Cohn, A. C. (1998): The phonetics-phonology interface revisited: where’s phonetics? Texas Linguistic Forum 41: 25–40.

Cohn, A. C. (2007): Phonetics in phonology and phonology in phonetics. Working Papers of the Cornell Phonetics Lab 16: 1–31.

DiCanio, C. T. (2008): The phonetics and phonology of San Martín Itunyoso Trique. Ph.D. diss. University of California, Berkeley.

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 in Linguistics 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 medical research methodology, 13(1): 1-15.

Fujisaki, H. (2003): Prosody, information, and modelling: With emphasis of Tonal Features of Speech. Proc. Workshop on Spoken Language Processing, 5-14. Mumbai.

Fulop, S. (2011): Speech Spectrum Analysis. Springer, Berlin.

Gandour, J. (1974): Consonant types and tone in Siamese. J. Phonet. 2: 337-50.

Gibbons, R. D.; Hedeker, D.; DuToit, S. (2010): Advances in analysis of longitudinal data. Annual review of clinical psychology, 6, 79-107.

Glover, W. (1970): Gurung tone and higher levels. Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics 3: 52-73.

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

Hale, M.; C. Reiss. (2000): Substance abuse and dysfunctionalism: Current trends in phonology. Ling. Inq. 31:157-169.

Hanson, H. M. (1995): Glottal characteristics of female speakers. Ph.D. diss. Harvard University, MA.

Hanson, H. M. (2009): Effects of obstruent consonants on fundamental frequency at vowel onset in English. J. acoust. Soc. Am. 125(1): 425-441.

Haraguchi, S. (1977): The Tone Pattern of Japanese: An Autosegmental Theory of Tonology. Tokyo: Kaitakusha.

Hastie, T.; Tibshirani, R. (1986): Generalized Additive Models. Statistical Science (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.

Haudricourt, A. G. (1965): Les mutations consonantiques des occlusives initiales en mon-khmer. Bulletin de la Société de Linguistique de Paris 60: 160-172.

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. PhD. diss. Berkeley: University of California.

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., Tone: A Linguistic Survey, pp. 77-111. Academic Press, New York.

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.

Hyman, L. M. (1973a): Consonant types and tone. Southern California occasional papers in linguistics No. 1. USC, Los Angeles.

Hyman, L. M. (1973b): The role of consonant types in natural tonal assimilations. In Southern California occasional papers in linguistics 1: 151-79.

Hyman, L. M. (1975): Phonology: theory and analysis. Holt, Rinehart and Winston, New York.

Hyman, L. M. (1976): Phonologisation. In A. Juilland, Linguistic studies presented to Joseph H. Greenberg, pp. 407–18. Anna Libri, Saratoga, Calif.

Hyman, L. M. (2013): Enlarging the scope of phonologisation. Origins of sound change: Approaches to phonologisation. Oxford University Press, UK.

Hyman, L. M.; Schuh, R.G. (1974): Universals of tone rules: evidence from West Africa. Ling. Inq. 5: 81-115

Ishihara, S. (1998). Independence between consonantal voicing and vocoid F0 perturbation in English and Japanese. In R. Manell and J. Robert-Ribes (eds.), Proceedings of the 5th International Conference of Spoken Language and Processing, pp. 3107-3110.

Jangjamras, J. (2012): Perception and production of English lexical stress by Thai speakers. PhD. diss. University of Florida.

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. (1996): The phonology-phonetics interface. In U. Kleinhenz, Interfaces in phonology, pp. 262–78. Akademie Verlag, Berlin.

Keyser, S. J.; Stevens, K. N. (2001): Enhancement revisited. In M. Kenstowicz, K. Hale: A life in language, pp. 271–91. MIT Press, Cambridge, MA.

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 P. de Lacy, The Cambridge handbook of phonology, pp. 401–34. Cambridge University Press, Cambridge.

Kirby, J. (2013): The role of probabilistic enhancement in phonologisation. Origins of Sound Change: Approaches to Phonologisation, pp. 228-246. Oxford University Press, Oxford.

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 M. V. Oostendorp, C. Ewen, E. Hume and K. Rice (eds.) The Blackwell Companion to Phonology, 5, pp. 2879-2907. Malden, MA and Oxford: Wiley-Blackwell.

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.

Li, R.; Zhang, H. Y. (1993): Chongming Fangyan Zidain (A dictionary of Chongming Dialect). Jiangsu Education Publishing House.

Liberman, A. M.; Delattre, P. C.; Gerstman, L. J.; Cooper, F. S. (1956): Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. Exp. Psychol. 52(2): 127-137.

Lisker; Abramson. (1964): A cross-language study of voicing in initial stops: Acoustical measurements. Voice onset time (VOT). Word 20: 384-422.

Ma, Y.; Mazumdar, M.; Memtsoudis, S. G. (2012): Beyond Repeated measures ANOVA: advanced statistical methods for the analysis of longitudinal data in anesthesia research. Regional anesthesia and pain medicine 37(1): 99-105.

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.

Mazaudon, M. (2012): Paths to tone in the Tamang branch of Tibeto-Birman (Nepal). In D. V. Gunther, Guido Seiler, The dialect laboratory: dialects as a testing ground for theories of language change, pp. 139-178. Studies in Language Companion Series 128. John Benjamins, Amsterdam.

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 L.M Hyman, (ed.). Studies in Stress and Accent [Southern California Occasional Papers in Linguistics 4]: 262-272.

Mei, T. (1970): Tones and prosody in Middle Chinese and the origin of the rising tone. Harvard Journal Asiatic Studies 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. The FASEB journal, 1(5), 365-374.

Nolan, F. (2003): Intonational equivalence: An experimental evaluation of pitch scales. Proc. the 15th Int. Congress of Phonet. Sci, 771-774. Barcelona.

Ning, L. H.; Shih, C.; Loucks, T. M. (2014). Mandarin tone learning in L2 adults: A test of perceptual and sensorimotor contributions. Speech Communication, 63-64, 55-69.

Ohala, J. J. (1974): A mathematical model of speech aerodynamics. Proc. of the Speech Communication Seminar 2: 65-72. Stockholm.

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. 3rd Int. Conf. on Spoken Lang. Processing (ICSLP 94), Yokohama, Japan, September 18–22. 491–494.

Pham, A. H. (2003): Vietnamese tone: A new analysis. Outstanding Studies in Linguistics. Routledge, London.

Potter, R. K.; Kopp G. A.; Green H. C. (1947): Visible speech. New York: D. Van Nostrand.

Prince, A.; Smolensky, P. (1993): Optimality Theory: constraint interaction in generative grammar. John Wiley & Sons.

Przezdziecki, M. A. (2005): Vowel harmony and coarticulation in three dialects of Yoruba: phonetics determining phonology. PhD. diss. Univ. Ithaca.

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.

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. Vancouver, UBC press.

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.

R Core Team (2013): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

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.

Shih, C.; Lu, H. Y. D. (2015): Effects of talker-to-listener distance on tone. Journal of 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 R. Togneri, 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 M. Broe, J. Pierrehumbert, Papers in laboratory phonology V: Acquisition and the lexicon, pp. 313 – 34. Cambridge University Press, Cambridge.

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.

Thurgood, G. (2002): Vietnamese tone: revising the model and the analysis. Diachronica 19: 333-363.

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.

Tsuchida, A. (1997): Phonetics and Phonology of Japanese Vowel Devoicing. PhD diss., Cornell University.

Ullah, S.; Finch, C. F. (2013): Applications of functional data analysis: A systematic review. BMC medical research methodology 13(1): 1.

Wayland, R.; Jongman, A. (2003): Acoustic correlates of breathy and clear vowels: the case of Khmer. J. Phonet. 31: 181–201

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. Proc. Speech Prosody. Dresden, Germany.

Wood. S. (2006): Generalized Additive Models: An Introduction with R. Taylor and Francis Group, LLC.

Xu, C. X.; Xu, Y. (2003): Effects of consonant aspiration on Mandarin tones. J. Int. Phonet. Ass. 33: 165–181.

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.; Prom-on, S. (2014): Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning. Speech Communication, 57, 181 – 208.

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.; 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.

Zhang, H. Y. (2009): Chongming Fangyan Yanjiu (The Study of Chongming Chinese). Beijing: China Social Sciences Press.

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.

Zhu, X. S. (1999): Shanghai Tonetics. Lincom Europa, Newcastle.

Zsiga, E. C. (1995): An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In B. Connell, A. Arvaniti, Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV, pp. 282-302. Cambridge University Press, Cambridge.