Tonal Aspects of Languages 2016
24-27 May 2016, Buffalo, New York

Complete and incomplete neutralisation in Fuzhou tone sandhi

Yang Li

Phonetics Laboratory, DTAL, University of Cambridge, United Kingdom
yl492@cam.ac.uk

Abstract

This is a study of incomplete neutralisation using Fuzhou tone sandhi as a test case. In Fuzhou Min, Tone 44 and Tone 232 undergo putative neutralisation into Tone 53 preceding a set of low tones, while T44 and T53 are neutralised into T44 preceding T53. Data from 10 Fuzhou speakers show the former neutralisation to be acoustically incomplete, with the sandhi tone of underlying Tone 44 having a higher pitch onset, while the latter appears to be acoustically complete. This result expands the typology of incomplete neutralisation by showing that two distinct phonological objects can incompletely neutralise (A / B → C) into a third object, as previous studies predominantly feature neutralisation of one object into another: A / B → B, e.g. Beijing tone sandhi T2/T3 → T2 / __ T3. It is argued that a hybrid model with both abstract categories and stored exemplars can best account for the data (Pierrehumbert, 2002 & 2006).

Index Terms: tone sandhi, incomplete neutralization, Fuzhou Min, hybrid model

1. Introduction

1.1. Fuzhou tone sandhi

Fuzhou is a Min Chinese dialect mainly spoken in Fuzhou, Fujian, China. Fuzhou has seven citation tones: five full tones and two checked tones (syllables with final glottal stops).

Table 1. *Fuzhou tone sandhi, adapted from [1].*

<table>
<tr><td>Context tone →</td><td>44</td><td>53</td><td>32</td><td>212</td><td>232</td></tr>
<tr><td>Target tone ↓</td><td colspan="5">Sandhi tone ↓</td></tr>
<tr><td>44</td><td colspan="2" rowspan="4">44</td><td colspan="3" rowspan="3">53</td></tr>
<tr><td>213</td></tr>
<tr><td>242</td></tr>
<tr><td>53</td><td>31</td><td colspan="2">21</td></tr>
<tr><td>31</td><td colspan="2">21</td><td>24</td><td colspan="2">44</td></tr>
</table>

Typologically, Fuzhou tone sandhi falls into the class of right-dominant sandhi systems [2]. In a multi-syllable tonal domain, the final syllable retains its citation tone, while all pre-final (target) syllables undergo sandhi. The identity of the resulting (surface/sandhi) tones is determined jointly by the identity of the target tone and the context tone. Table 1 is an illustration: the top row shows the context tone, the leftmost column shows target tone, and cells in the middle show the realisations of target tones before respective context tones. At least on the surface, Fuzhou tone sandhi appears to involve the paradigmatic replacement of whole tones [3], [4]; it then constitutes a classic example of phonological neutralisation.

Two neutralisation groups will be the focus in this paper. They are listed in (1) and (2) and are in bold in Table 1.

(1) 44, 53 → 44 / ___ 53
(2) 242, 44 → 53 / ___32

1.2. Incomplete neutralisation

I take incomplete neutralisation (IN) to mean phonological neutralisations that produce perceptually (almost) equivalent outputs that nevertheless leave instrumentally detectable differences. IN has been reported for both segmental and tonal neutralisations, especially in final devoicing processes in Germanic and Slavic languages and Mandarin tone sandhi, e.g. [5], [6].

Theoretically, IN is a challenge for generative phonology [7]. The existence of incompletely neutralised objects (e.g. coda in German *rad*), which are differentiated from another class of objects (e.g. coda in German *rat*) in production but not in perception is hard to accommodate in a classical modular feed-forward model of phonology [8]. Crucially, if category change (e.g. from underlying voiced to surface voiceless) must be discrete and must take place in the phonological component, and if phonetics is not allowed to “look back” to lexical and morphological information, then phonologically neutralised objects should have identical acoustic outputs. IN is then unexpected for the classical model.

With considerable theoretical stakes, debate on the reliability of IN has continued for decades. One key confounding factor in final devoicing IN is orthography [7], [9]–[11], e.g. German *rad* and *rat* suggesting underlying voicing. This is much less of concern in tone sandhi. Fuzhou, for example, is a largely unwritten language; when it is written, the Chinese script used for this purpose does not mark tones, so potential orthographic effect is kept to a minimum.

Investigating Fuzhou tone sandhi can also expand the typology of IN in an empirically-enriching and theoretically informative way. Most neutralisations investigated to date are of the A / B → B type, e.g. /d/ and /t/ neutralised into /t/. If we make the rather trivial observation that the output of a derivation (assuming there is one) is inherently more variable than an “unaltered” object, then this is a potential confound inherent in IN studies. Thanks to the scale of tonal neutralisation in Fuzhou, it is possible to study a neutralisation into a surface form which is distinct from all the underlying tones, e.g. in (2).

1.3. Previous studies on tone sandhi IN

Multiple studies have reported that in Beijing Mandarin Tone 3 sandhi [6], [12]–[14], surface T24 resulting from tone sandhi is lower in pitch than lexical T24, with varying reported mean difference: 20 Hz [13] to 3.2 Hz [13]. On the other hand, two studies on Taiwanese Southern Min and Taiwanese Mandarin have not turned up statistically significant difference between sandhi and lexical tones [16], [17]. On Fuzhou, [1] fails to find an acoustic difference along the duration and mean F0 dimensions for Fuzhou sandhi T44 and lexical T44. The goal of this study is to partially replicate and expand on [1] by investigating further neutralisation groups, comparing the whole contours of F0 realisation, and using statistical tools that can evaluate both complete and incomplete neutralisation (see section 2.3).

http://dx.doi.org/10.21437/TAL.2016-25

2. Method

2.1. Speakers

10 native speakers of Fuzhou (aged from 30 to 45, mean = 38.5) completed the experiment out of the 13 recruited. All are residents in the city of Fuzhou, speak Mandarin as L2 and can read and write in Chinese. All have completed at least high school education and none reported hearing problems. Before the experiment, speakers were shown 15 common words in simplified Chinese on a computer screen, drawn from [18]. To pass, speakers had to correctly pronounce the Fuzhou reading of 14 words within 7 seconds of reaction time. 3 speakers therefore did not participate in the experiment.

2.2. Stimuli and procedure

The stimuli are disyllabic words embedded in an invariant carrier phrase [ŋuai^{31} puoʔ^{44} t^{h}øy^{44} ___ k^{h}øy^{44} ny^{31} t^{h}iaŋ^{55}], "I want to read ____ for you to listen". Four sets of words were selected, two for each neutralisation group in (1) and (2). Within the neutralisation group, the words will crucially differ in the first member of the underlying tone sequences. The null hypothesis is that despite this difference, the resulting sandhi tone should be identical, as in Table 2. The test words are controlled for segments on the first syllable between minimal underlying tone pairs.

Table 2. *Stimuli design: the underlying difference in the first tone is supposedly neutralised.*

UR Tone pairs	Neutralisation into
/T53.T53/ /T44.T53/	[T44.T53]
/T232.T32/ /T44.T32/	[T53.T32]

During the experiments, the screen would display the test words in large fonts in randomised order. Participants were instructed to pronounce the words in isolation in a clear manner, followed by a pause, and then a carrier sentence in which the test words were embedded. This was recorded with a Nagra ARES-M II recorder held 25 cm away from the speakers. The total tokens obtained are 2560 (2 neutralisation groups * 2 tone pairs within a group * 8 segmental types * 4 repetitions * 10 speakers * 2 speech rate conditions). Only results from carrier sentences are reported.

2.3. Data processing and analysis

F0 data were extracted with VoiceSauce [19] and time-normalised by taking 15 equidistant measurements through the rhyme. These are transformed into semitones [20]. Duration of the rhyme/syllable was also obtained, with the twofold goals of detecting IN along the duration dimension (e.g. Cantonese in [21]) and validating the timing-normalisation of F0 contour. This is mindful of the fact that, if Fuzhou adopts a "compression" strategy [22] for realising F0 on varying sonorant material, the shape of the F0 contour will be directly influenced by duration, which makes time-normalised F0 contour potentially misleading.

Statistical analysis combines Smoothing-Spline ANOVA (SS ANOVA) [23] and Bayesian inference [24]. SS ANOVA fits polynomial functions on discrete data points, creating smoothed splines with surrounding ribbons that represent the 95% confidence intervals estimated from the data. The non-overlap of two 95% CI ribbons will signify a statistically significant difference. The analysis here is conducted in *R* with the "gss" package.

The adoption of Bayesian inference addresses the bias in IN literature unfairly favouring the alternative hypothesis. In traditionally used "frequentist" statistics framework, a small *p* indicates statistical significance, but a large *p* is uninformative - the absence of evidence does not equate evidence of absence. Bayesian inference corrects this problem by calculating the Bayes Factors of alternative hypotheses, evaluating their odds against the null hypotheses. This has the double advantage of being able to both confirm and reject the null and provide graded evidence, thereby facilitating cross-study comparison and the accumulation of evidence. The analysis is conducted with the "BayesFactor" package in *R* using [24]'s procedure.

3. Results

3.1. Complete neutralisation into T44

The average duration and estimated confidence interval of the sandhi tone syllables are listed in Table 3. No significant difference as a function of underlying tones was found (by Linear mixed effect models, p= .32), with the 95% confidence intervals almost completely overlapping.

Table 3. *The mean and 95% confidence interval of the duration of the surface T44.*

Tone pairs	Mean duration (ms)	95% CI (ms)
/T53.T53/	137.01	[132.72, 141.85]
/T44.T53/	136.44	[133.21, 140.13]

The mean F0 contour of surface T44 is depicted in Figure 1 through the thick red (sandhied T44) and brown lines (sandhied T53). The thinner lines surrounding these enclose the 95% confidence intervals of the respective contours as estimated through SS ANOVA. It can be seen that both the mean F0 as well as the confidence intervals largely overlap with each other, suggestive of complete neutralisation.

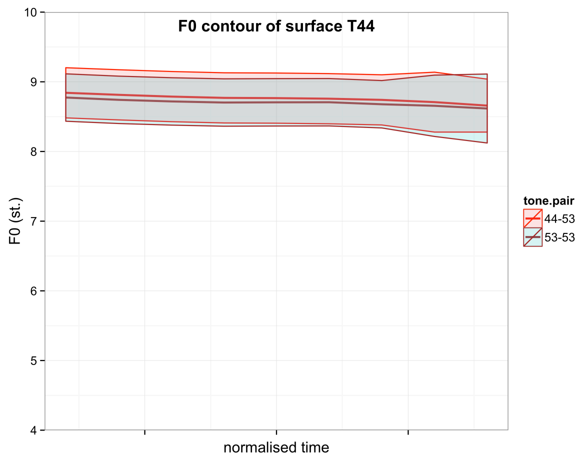

Figure 1: *Time normalised F0 contour of sandhi T44 from underlying T53 and T44, in semitones.*

The mean difference in F0 between surface T44 stemming from underlying T44 and T53 is then individualised and plotted in Figure 2. We count 4 speakers whose mean F0 of sandhied T44 is lower than sandhied T53 and 6 speakers exhibiting the reverse pattern. The size of the effect is generally small, and no difference more than 0.5 semitones between the underlying tone pairs is found for any individual speaker. Hence no clear effect in either direction emerges.

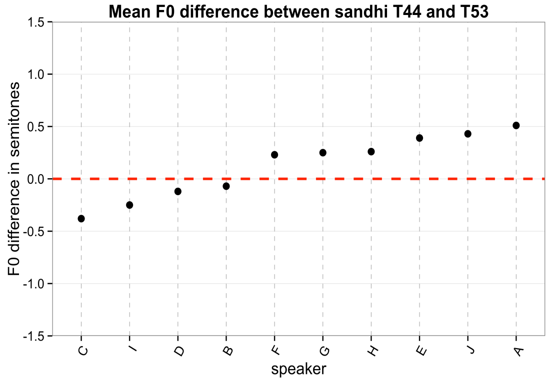

Figure 2: *Mean overall F0 difference in sandhi tone from T44 and T53, averaged over each of 10 speakers.*

Finally, we calculate Bayes Factors of various alternative hypotheses based on the observed data. The Bayes Factor of a given hypothesis is the odds of the hypothesis being true relative to the null hypothesis. The Bayes Factors for incomplete neutralisation into T44 is 1.19, indicating that it is about as likely as the null hypothesis, and that the difference is "not worth more than a bare mention", according to the criteria in [25]. The combined evidence from SS ANOVA, individual variation and Bayesian inference confirm that the neutralisation into T44 is acoustically complete.

3.2. Incomplete neutralisation into T53

The average duration and estimated confidence interval of the sandhi tone syllables are listed in Table 3. No significant difference as a function of underlying tones was found (by linear mixed effects models, p= .32), with the 95% confidence intervals almost completely overlapping.

Table 4. The mean and 95% confidence interval of the duration of the surface T53.

Tone pairs	Mean duration (ms)	95% CI (ms)
/T232.T32/	126.30	[121.72, 130.85]
/T44.T32/	124.83	[121.27, 128.19]

The mean F0 contour of surface T53 is depicted in Figure 3. It can be seen that in both underlying tone conditions, the pitch of the sandhied tones starts from the top of the pitch range, falls hesitantly at the beginning then travels more steeply downward, reflecting the citation acoustics of T53. While the 95% confidence interval regions of the two tone pairs overlap with each other on most of the time points, the initial section (0~20%) of the F0 from underlying T44 is significantly higher than that of underlying T232. The average difference is 0.68 semitones, or 6.75 Hz. As can be further gleaned from Figure 3, this difference persists until about 70% into the rhyme, when sandhied T44 begins to be lower than sandhied T232. Sandhi T44 thus starts with higher F0 and terminates with lower F0, with an average falling range of 4.75 semitones compared to the 3.89 semitones of sandhi T232. This difference is statistically significant (p< .001).

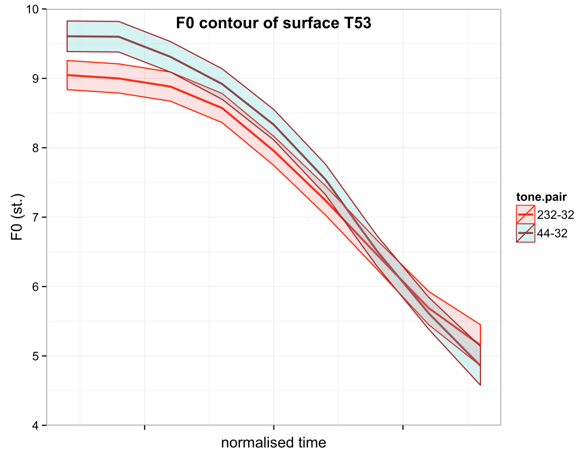

Figure 3: *Time normalised F0 contour of sandhi T53 from underlying T44 and T232, in semitones.*

Figure 4 shows the pattern of individual differences. Note that the scale on y-axis in Figure 4 represents considerably larger differences between underlying tonal categories than in Figure 2. As shown in Figure 4, eight speakers have higher overall F0 in sandhi T44, and only one outlier ("B") trends in the opposite direction. The absolute size of the differences for most speakers is between 0.5 and 1 semitones. This is above the just-noticeable threshold for pitch and thus physiologically perceptible [26], but probably not enough for phonological contrast [27]; in other words, within the range where IN can be expected. The consistency of individual realisation and the appropriate size of the f0 difference provide further evidence that surface T53 from the sandhied T44 does have higher onsets than that from sandhied T232.

Bayesian inference is then conducted to validate the IN effect in neutralisation into T53. The alternative hypothesis consisting of underlying tone pair as fixed effects and speaker as random effects has a Bayes Factor of 13.26, i.e. it is 13.26 times more likely to be true than the null. This constitutes "positive" evidence that sandhied T44 differs from sandhied T232 [25]. Based on combined evidence, I reach the tentative conclusion that sandhi T44 and T232 are largely, but incompletely neutralised into T53.

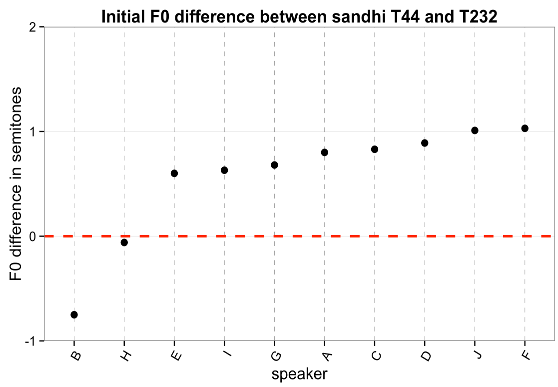

Figure 4: *Mean overall F0 difference in sandhi tone from T44 and T232, averaged over each of 10 speakers.*

4. Discussion

The results presented above have been obtained through the combined examination of 95% confidence intervals from SS ANOVA, individual variation and Bayes Factors. The overall picture to emerge from these limited data is that we can be fairly confident of complete neutralisation in T53/T44 → T44 and incomplete neutralisation in T232/44 → T53. The IN effect in the latter case is the second reported case of IN in tone sandhi besides Beijing Mandarin. This further confirms tone sandhi as a fertile ground for studying IN, with its resistance to many confounds present in segmental studies, e.g. the influence of orthography hinting an underlying contrast. Nevertheless, it is prudent to interpret the present finding of IN in view of other potential confounds such as lexical frequency [28], especially due to the relatively small size of difference inherent in IN studies.

The finding of IN in Fuzhou T232 & T44 → T53 is also the first sighting of "A & B → C" type of incomplete neutralisation. This class of IN is free from the logical confound of comparing a phonological object which is inherently variable (e.g. Mandarin lexical T3 must satisfy a set of prosodic condition to become surface T2) against a phonological object inherently more stable (Mandarin lexical T2). This puts IN on a firmer empirical footing, given the controversies which have surrounded it since its inception.

The use of Bayesian inference confirms the complete neutralisation of T53 and T44 into T44, a conclusion which traditional methods are in principle unable to draw. I argue that it is both worthwhile and feasible to introduce Bayesian inference into IN studies. Worthwhile because of its ability to both confirm and reject IN and the graded nature of its conclusions; feasible because recent advances are making Bayesian hierarchical modelling and inference increasingly user-friendly [29].

4.1. Theoretical implications

While the existence of IN is far from a settled question, there have been a number of proposed accounts of IN. A broad dividing line can be made: Some segregate phonetics from pre-phonological modules (e.g. lexicon, morphology) and attribute the observed acoustic difference to a formal, phonological difference. Others will allow phonetics to directly access lexical and morphological information, in doing so relaxing the modular feed-forward assumption [8]. Much of the discussion has centred on final devoicing, and tone sandhi data provided here can be illuminating.

One theory that deals with final devoicing IN is Turbidity Theory [30]. It is assumed that underlyingly voiced obstruents are specified with a monovalent [voice] feature, absent in underlying voiceless obstruents. Highly ranked FINDEV constraint forces [voice] to be unparsed, and this can be interpreted by phonetics "in a number of very subtle ways" [30, p. 1371], one of which is IN. A similar strategy is employed by [31], where sonorants and voiced obstruents share a SV (spontaneous voicing) node in the feature geometry, which voiceless obstruents lack. IN is argued to follow from the presence of SV node in voiced obstruents when the feature difference on the laryngeal node is neutralised by final devoicing. The discovery of both complete and incomplete neutralisation within the same system poses a challenge. The tonal analogue to the supposed difference in segmental [t] and [d̥] is hard to conceive, especially when tone sandhi here seems to involve whole-tone substitution rather than feature alteration. Further, theories along this line have to posit differences in phonological computation only for those tone sandhi groups which show IN effect, but not for sandhi groups which do not, reducing predictive power.

The solution to IN favoured here is a hybrid model of phonological competence which encompasses both abstract categories and stored exemplars [32]–[34]. One such account is co-activation of morphologically related forms [35]: realisations of words or stems in some contexts may influence its realisations in others. In the present example, the otherwise low starting pitch of Fuzhou T232 might have subtly nudged its surface T53 variant towards a low pitch onset. Assuming that exemplar storage within our hybrid model takes place both at the morpheme and the word level (a similar line of thinking can be found in [36]), we can compare morphemes from two tonal classes: the exemplar clouds for T44 will include the realisations of both surface T53 and T44 across all phonological contexts, while those of T232 will include surface T53 and T232. It follows that surface T53 drawn from all activated tokens of lexical T44 are more likely to resemble T44 (with higher pitch onset), and surface T53 from all instances of lexical T232 is biased to behave comparatively more like T232 (with lower pitch onset). A validation of this account, however, will take far more than an acoustic study of Fuzhou tone sandhi to achieve.

5. Conclusions

Using acoustic data from 10 Fuzhou speakers under laboratory condition, the present paper reports both complete and incomplete neutralisation in Fuzhou tone sandhi. This finding represents a non-trivial expansion of the typology of incomplete neutralisation and demonstrates that it may occur in neutralisation of the "A & B → C" type.

The result is argued to show the vulnerability of accounts of incomplete neutralisation which attribute it to a formal difference in the phonological component [30], [31]. I argue for a hybrid model of phonology which utilises both abstract categories and stored exemplars, as it seems more consistent with results from other studies of IN [10], [36], [37] and the present contribution. Minimally, it seems necessary to allow for direct interaction between phonetics and the morphology and the lexicon.

6. Acknowledgements

I thank Francis Nolan, Bert Vaux, Brechtje Post and Adrian Leemann for discussions. Thanks are also due to audiences at ICPhS 2015 at Glasgow and at Experimental Phonology and Phonetics Seminar at Cambridge. Reviews from the Scientific Committee of the 5th International Symposium on Tonal Aspects of Languages greatly improved this paper; all remaining errors are mine. Financial support comes from the External Research Studentship of Trinity College, Cambridge, and is gratefully acknowledged.

7. References

[1] Y. Li, "Tone sandhi and tonal coarticulation in Fuzhou Min," in *Proceedings of ICPhS 2015*, 2015.

[2] A. Leemann, M.-J. Kolly, Y. Li, R. Chan, G. Kwek, and A. Jespersen, "Towards a typology of prominence perception: the role of duration," in *Proceedings of Speech Prosody 2016*, 2016.

[3] Y. Li, "An experimental study of Fuzhou tone sandhi," Unpublished MPhil thesis, University of Cambridge, Cambridge, 2014.

[4] R. Schuh, "Tone rules," in *Tone: a linguistic survey*, V. A. Fromkin, Ed. San Diego, CA: Academic Press, 1978, pp. 221–256.

[5] R. Port and M. O'Dell, "Neutralization of syllable-final voicing in German," *Journal of Phonetics*, vol. 13, no. 4, pp. 455–471, 1985.
[6] S. H. Peng, "Lexical versus 'phonological' representations of Mandarin sandhi tones," in *Papers in laboratory phonology 5: acquisition and the lexicon*, M. B. Broe and J. Pierrehumbert, Eds. Cambridge: Cambridge University Press, 2000, pp. 152–167.
[7] A. Manaster Ramer, "A letter from an incompletely neutral phonologist," *Journal of Phonetics*, vol. 24, no. 4, pp. 477–489, Oct. 1996.
[8] R. Bermúdez-Otero, "Diachronic phonology," in *The Cambridge handbook of phonology*, P. De Lacy, Ed. 2007, pp. 497–517.
[9] N. Warner, E. Good, A. Jongman, and J. Sereno, "Orthographic vs. morphological incomplete neutralization effects," *Journal of Phonetics*, vol. 34, no. 2, pp. 285–293, 2006.
[10] M. Ernestus, "Gradience and categoricality in phonological theory," in *The Blackwell companion to phonology.*, Wiley-Blackwell, 2011.
[11] D. A. Dinnsen and J. Charles-Luce, "Phonological neutralization, phonetic implementation and individual differences," *Journal of Phonetics*, vol. 12, no. 1, pp. 49–60, 1984.
[12] P. Kratochvil, "The case of the third tone," in *Wang Li Memorial Volumes: English Volume*, Chinese Language Society of Hong Kong, Ed. Hong Kong: Chinese University Press, 1987, pp. 253–277.
[13] E. Zee, "A spectrographic investigation of Mandarin Tone Sandhi," *University of California Working Papers in Phonetics*, vol. 49, pp. 98–116, Apr. 1980.
[14] J. Yuan and Y. Chen, "3rd tone sandhi in standard Chinese: A corpus approach," *Journal of Chinese Linguistics*, vol. 42, no. 1, 2015.
[15] Y. Xu, "Contextual tonal variation in Mandarin Chinese," Doctoral dissertation, University of Connecticut, Storrs, 1993.
[16] J. Myers and J. Tsay, "Investigating the phonetics of Mandarin tone sandhi," *Taiwan Journal of Linguistics*, vol. 1, no. 1, pp. 29–68, 2003.
[17] J. Myers and J. Tsay, "Neutralization in Taiwan Southern Min tone sandhi," in *Interfaces in Chinese Phonology: Festschrift in honor of Matthew Y. Chen on his 70th birthday*, Y. E. Hsiao, H. Hsu, L.-H. Wee, and D. Ho, Eds. Taipei: Academia Sinica/Institute of Linguistics, 2008, pp. 47–78.
[18] R. Li, Y. Liang, G. Zou, and Z. Chen, *Fuzhou fangyan cidian [A dictionary of Fuzhou Dialect]*. Fuzhou, China: Fujian Renmin Chubanshe, 1994.
[19] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, "VoiceSauce: a program for voice analysis," in *Proceedings ICPhS 2011*, 2011, pp. 1846–1849.
[20] F. Nolan, "Intonational equivalence: an experimental evaluation of pitch scales," in *Proceedings of the 15th international congress of phonetic sciences*, Barcelona, 2003, pp. 771–774.
[21] A. C. L. Yu, "Understanding near mergers: the case of morphological tone in Cantonese," *Phonology*, vol. 24, no. 1, pp. 187–214, Jan. 2007.
[22] E. Grabe, B. Post, F. Nolan, and K. Farrar, "Pitch accent realization in four varieties of British English," *Journal of Phonetics*, vol. 28, no. 2, pp. 161–185, Apr. 2000.
[23] C. Gu, *Smoothing Spline ANOVA Models*. Springer Science & Business Media, 2013.
[24] J. Rouder, R. D. Morey, J. Verhagen, A. R. Swagman, and E.-J. Wagenmakers, "Bayesian analysis of factorial designs," Submitted.
[25] R. E. Kass and A. E. Raftery, "Bayes factors," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 773–795, Jun. 1995.
[26] C. Liu, "Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners," *The Journal of the Acoustical Society of America*, vol. 134, no. 4, pp. 3011–3020, 2013.
[27] D. Silverman, "Pitch discrimination during breathy versus modal phonation," in *Phonetic Interpretation Papers in Laboratory Phonology VI*, J. Local, R. Ogden, and R. Temple, Eds. Cambridge: Cambridge University Press, 2004, pp. 293–304.
[28] Y. Zhao and D. Jurafsky, "The effect of lexical frequency and Lombard reflex on tone hyperarticulation," *Journal of Phonetics*, vol. 37, no. 2, pp. 231–247, 2009.
[29] T. Sorensen and S. Vasishth, "Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists," *arXiv:1506.06201 [stat]*, Jun. 2015.
[30] M. van Oostendorp, "Incomplete devoicing in formal phonology," *Lingua*, vol. 118, no. 9, pp. 1362–1374, Sep. 2008.
[31] K. Rice and P. Avery, "On the interaction between sonorancy and voicing," *Toronto Working Papers in Linguistics*, vol. 10, 1989.
[32] M. Ernestus, "Acoustic reduction and the roles of abstractions and exemplars in speech processing," *Lingua*, vol. 142, pp. 27–41, Apr. 2014.
[33] J. Pierrehumbert, "Word-specific phonetics," *Laboratory phonology*, vol. 7, pp. 101–139, 2002.
[34] J. Pierrehumbert, "The next toolkit," *Journal of Phonetics*, vol. 34, no. 4, pp. 516–530, Oct. 2006.
[35] M. Ernestus and R. H. Baayen, "The functionality of incomplete neutralization in Dutch: The case of past-tense formation," *Laboratory phonology*, vol. 8, no. 1, pp. 27–49, 2006.
[36] A. Braver, "Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping," *Lingua*, vol. 152, pp. 24–44, Dec. 2014.
[37] T. B. Roettger, B. Winter, S. Grawunder, J. Kirby, and M. Grice, "Assessing incomplete neutralization of final devoicing in German," *Journal of Phonetics*, vol. 43, pp. 11–25, Mar. 2014.