



Surface vs. Underlying Listening Strategies for Cross-Language Listeners in the Perception of Sandhied Tones in the Nanjing Dialect

Xin Li¹, René Kager¹, Wentao Gu²

¹Utrecht Institute of Linguistics, Utrecht University, The Netherlands

²Nanjing Normal University, China

xin.li@uu.nl, R.W.J.Kager@uu.nl, wtgu@njnu.edu.cn

Abstract

This study is devoted to exploring surface/underlying listening strategies adopted by native and non-native, tone and non-tone language groups in their perception of sandhied tones, and the possible effect of the coarticulatory/non-coarticulatory nature of the sandhi rule on tone perception. The mapping between surface sandhied tones and underlying tones is investigated by a pair of Nanjing sandhi rules, one coarticulatory and the other non-coarticulatory, involving three groups of listeners: Dutch, Beijing, and native Nanjing, by means of a Concept Formation paradigm. Results reveal distinct perceptual patterns in the three groups. Dutch listeners experience difficulty in creating phonological representation for tones even in surface listening; Beijing listeners use their native ability at tone perception for interpreting Nanjing sandhied tones in surface tone perception, but have no access to underlying patterns; Nanjing listeners perceive native sandhied tones at both the underlying level and surface level and seem to always mix them when analyzing tones. The coarticulatory/non-coarticulatory nature of sandhi rules does not seem to play a role in the current experiment.

Index Terms: surface/underlying listening strategy, coarticulation/non-coarticulation, perception of sandhied tone, cross-linguistic tone perception

1. Introduction

Tone sandhi is generally known as the phenomenon of underlying lexical tones being modified under the influence of their tonal context and surfacing with different phonetic forms [1]. It is widely found in various Chinese dialects. For example, Beijing Mandarin is well-known to have a T3 sandhi which involves a lexical tone T3 changing to an actually articulated T2 when followed by another lexical T3. It can be written as $\underline{T3}T3 \rightarrow \underline{T2}T3$. (The underlined tones indicate the tones in change.) Considering that the sandhied tone originates from an underlying tone but surfaces possibly as another tone, the question arises how listeners categorize it.

It is hypothesized that two listening strategies will apply in processing sandhied tones: the surface-based strategy and the underlying-based strategy. The surface-based strategy refers to listening to the sandhied tone in a purely surface-oriented way focusing attention only on its surface phonetic forms, without aiming at its corresponding underlying base. The underlying-based strategy uses information about the mapping between a sandhied tone and its underlying counterpart based on native listeners' phonological knowledge of the sandhi rule.

Evidence for native listeners' underlying-based strategy was found in [2]. It was found that native listeners consistently categorized the sandhied tone in T3 sandhi of Beijing Mandarin as the underlying T3. However, the effects of the underlying tone were confounded with lexical knowledge in this experiment. Hence, it still remains a question whether

native listeners of a tone language that has a sandhi rule may perceive tone at the underlying level, and whether this hypothetical underlying listening strategy applies consistently in native listeners. In principle, the surface-based listening strategy could also work for natives.

The underlying-based strategy in perceiving sandhied tones should not be available to non-native listeners due to their lack of the phonological grammar of the target language. Regarding the surface-based strategy, non-natives' perception of tones is believed to be determined in large part by the degree to which non-native tones can be assimilated to native tonal or intonational category representations [3]. For non-tone-language listeners who lack contrastive categories of tones [4, 5], this perceptual assimilation is undoubtedly limited compared to tone-language listeners.

Yet there might be at least one circumstance under which non-natives might be able to access underlying elements, and the bridge that allows them to do so is their general (language-nonspecific) knowledge of coarticulation. Evidence for such perceptual ability was found in [6]'s observation that both place-of-articulation assimilation and voice assimilation rules allowed non-natives to relate the surface form to some extent back to the underlying pattern. This finding raises the question whether a coarticulation-based underlying listening strategy may also hold for tonal underlying-surface alternations; to be more specific, whether non-native listeners might have access to the underlying tone when a coarticulatory mapping can be established between the surface tone and an inferred underlying tone.

The goal of the current study is to compare the perception of sandhied tones targeting their surface or underlying tones in a coarticulatory sandhi rule and a non-coarticulatory sandhi rule, across three groups of listeners: a native group, a non-native tone-language group and a non-tone-language group.

A pair of coarticulatory and non-coarticulatory sandhi rules are found in the Nanjing dialect, a dialect in Hongchao dialectal area of Jianghuai Mandarin in China. This dialect has five lexical tones (a high-falling T1, a low-rising T2, a low-dipping T3, a high-level T4 and a high-arched T5) and several sandhi rules, documented in earlier studies [7, 8]. A production experiment conducted by the authors with 18 Nanjing speakers aged between 18 and 30 produced slightly different results from previous studies, and revealed two sandhi rules differing in their coarticulatory/non-coarticulatory nature. Figures 1 and 2 illustrate that in Sandhi 1 the offset of a high-falling T1 in the first syllable is raised to the same pitch as the onset of the following T1, which can be interpreted as a tonal coarticulation process; Sandhi 2 is a non-coarticulation process, since the offset of a high-level T4 in the first syllable deviates from the subsequent onset of a high-arched T5, instead of approaching it. It is noted that some speakers produce a steeper fall in the sandhied tone in Sandhi 2. Besides, the two sandhi rules both involve a Nanjing T1 vs. T4

contrast. The two sandhi rules are used as the base for creating stimuli for this study.

- Sandhi 1: T1T1 → T4T1
- Sandhi 2: T4T5 → T1T5

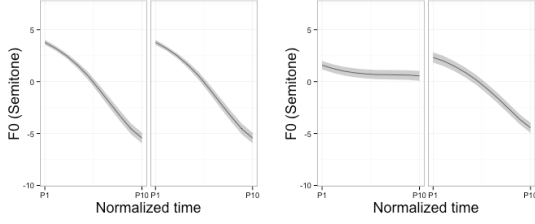


Figure 1: *Lexical T1+T1 (left) & sandhied T1+T1 (right) in the Nanjing dialect (averaged across 18 speakers). Solid lines indicate mean f_0 ; gray ribbons stand for ± 1 standard error of mean (also in Figure 2).*

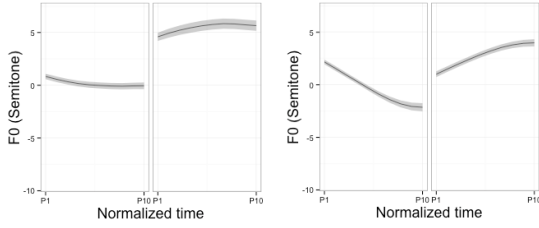


Figure 2: *Lexical T4+T5 (left) & sandhied T4+T5 (right) in the Nanjing dialect (averaged across 18 speakers).*

In order to test the underlying/surface listening strategies for sandhied tones in native and non-native listeners, the Concept Formation paradigm [9] was adopted as the paradigm in the current study. This paradigm can be used to reveal a category concept for listeners who already possess it as well as to create such a category concept for listeners who do not have it yet. It consists of a training session and a test session. In the training session, listeners are trained to form a category or make use of an existing category by listening to target tokens matching the target category and non-target tokens mismatching the category as well as receiving feedback for every token they hear. Subsequently, a test is given which includes (a) target tokens, (b) non-target tokens and (c) test tokens, which match both the target category and a different linguistic category and are thus ambiguous to the listeners who already possess the latter category. This paradigm is especially useful in bringing the latter category out.

It is expected that when presented with ambiguous tokens (sandhied tone) that match both the underlying and surface category (underlying tone and surface tone), native listeners, when using the underlying strategy, will consistently categorize the ambiguous tokens so as to match the underlying tone. So is the case if the surface strategy applies. Non-native listeners, who lack access to the underlying level but may possess a similar tonal category in their native phonology, will perceptually assimilate the target category to their native category at the surface level. Finally, non-native listeners who have no tonal category in their native phonology will create a new category perceiving the target category also at the surface level.

In the current study, we compare the mapping between the sandhied tones and their surface/underlying tones by exposing

half of the participants to tokens that invite them to learn the underlying tone as the target category, and exposing the other half to tokens that invite them to learn the surface tone. The experiment involved a Dutch group (henceforth, NL), a Beijing group (henceforth, BJ) and a native Nanjing group (henceforth, NJ).

2. Method

2.1. Stimuli

In the learning session, the target tokens were NJ disyllabic words beginning with the target tone followed by a variety of tones; non-target tokens were NJ disyllabic words beginning with other tones except the target tone. No sandhi was involved in the learning session. The number of target tokens (3 syllable structures * 8 tokens) and non-target tokens (8 structures * 3 tokens) were kept equal. Stimuli in the learning session were designed to allow participants to learn the target tone in the first syllable.

The test session included 8 target tokens, 8 non-target tokens and 16 test tokens. The target and non-target tokens in the test session were new, but constructed in the same way as those in the learning session. Again, no sandhi was involved in target or non-target tokens in the test session. Test tokens were Nanjing disyllabic words beginning with a sandhied tone, which theoretically matches both the target tone (underlying tone/surface tone) and its sandhi-related tone (surface tone/underlying tone).

Test tokens were split into a real-word and a non-word session, and were constructed in different ways to eliminate the influence from word familiarity and word-likeness. For test tokens in the real-word session, the disyllabic word bearing surface tones and its counterpart word bearing underlying tones were both existing words in the Nanjing lexicon. The frequency of each underlying-surface word pair was balanced using a survey of subjective familiarity ratings, in which 14 native Nanjing participants aged between 18 and 30 rated a written list of the words in random order on a scale ranging from 1 to 7 as the word occurs more often in their language. Only the words with approximately equal familiarity ratings in the underlying-surface pair were selected. For example, the words *xianT1 huaT1* ‘fresh flower’ and *xianT4 huaT1* ‘present flower to sb’ are both frequently used words in Nanjing. For test tokens in the non-word session, the disyllabic word bearing the surface tones and its counterpart word bearing the underlying tones are both gaps in the Nanjing lexicon. The word-likeness of each pair was controlled. For example, the words *shuT1 jiangT1* and *shuT4 jiangT1* are both meaningless in Nanjing.

All the tokens were spoken naturally by a 25-year-old female from Nanjing who produces a steeper fall as the sandhied tone in Sandhi 2. The target tokens and non-target tokens were produced without sandhi. The test tokens were produced in sandhi condition. All the tokens were examined in Praat [10] and the most typical ones of Sandhi 1 and Sandhi 2 patterns were selected as the auditory stimuli in the experiment.

2.2. Participants

Eighty participants with self-reported normal hearing were recruited separately for the NL, BJ and NJ group. All of them were aged from 18 to 30, which is consistent with the age span used in the previous production experiment and subjective familiarity/wordlikeness-rating task in Nanjing. None of the NJ listeners had taken part in the previous experiment. None

of the BJ listeners had been exposed to the Nanjing dialect before. Listeners in each language group were randomly assigned to Sandhi 1 and Sandhi 2, and then further randomly divided into two training conditions: (a) the underlying tone was the target category and (b) the surface tone was the target category. Every participant did both the real-word session and the non-word session in one experiment. The order of the two sessions was counterbalanced for participants.

2.3. Procedure

In the learning session, listeners were instructed to identify the melodic property which all the target tokens they would hear shared and which non-target tokens didn't. Each trial started with a stimulus followed by a 3-second silence, then auditory feedback was presented indicating membership of the stimulus in the target category (target/non-target). The feedback was in Dutch for NL listeners, in Beijing Mandarin for BJ listeners, and in the Nanjing dialect for NJ listeners to trigger each group's native tonal or intonational grammar. Participants had to make quick responses by pressing either the left or right shift key on the keyboard during the 3-second time window. Missed trials were repeated at the end of the trials.

The learning session automatically ended when a listener performed thirteen trials in a row with two or fewer errors, thus meeting the a priori criterion for having learned the target category. Then after a page of instructions they proceeded to the test session. In the test no auditory feedback was given.

The experiment was conducted via a ZEP program. Both the judgment (target/non-target) and reaction time for each test token were recorded. In the current paper only the judgment data are presented.

3. Results

A generalized linear mixed-effects model was constructed to analyze the judgement data. PARTICIPANT and ITEM were included as random effects. Significant variance was found in intercept across participants: $SD = 1.058$, $p < .05$. There was no significant variance in intercept across items: $SD = 0.064$, $p > .05$. Fixed effects of LANGUAGE GROUP, TRAINING CONDITION and the interaction between them were added to the model step by step and all proved to significantly improve the fit of the model. Yet adding SANDHI or any interaction between SANDHI and LANGUAGE GROUP/TRAINING CONDITION did not significantly improve the model fit. Table 1 shows the fixed effects of this model.

Figure 3-5 depict the probability of mapping the sandhied tone to the underlying/surface tone across conditions in each language group. Mapping in the two sandhi conditions are combined in figure 6.

Table 1. Estimated parameters of fixed effects in the mixed-effects model.

	B (SE)	95% CI for odds ratio		
		Lower	Odds ratio	Upper
Intercept	0.37(0.18)*			
BJ	-2.71 (0.27)***	0.04	0.07	0.11
NJ	0.77(0.25)**	1.31	2.16	3.56
Surface condition	-0.11(0.25)	0.55	0.90	1.47
BJ: Surface	5.41(0.39)***	104.38	224.27	481.87
NJ: Surface	0.12(0.36)	0.56	1.13	2.28

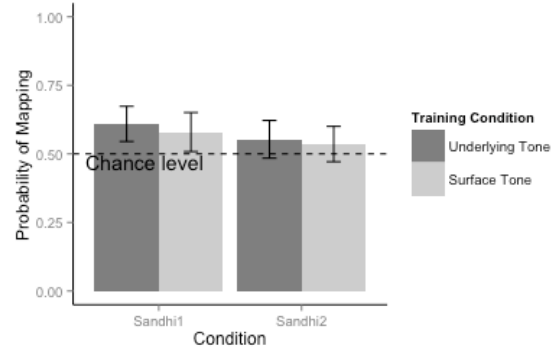


Figure 3: Probability of mapping across conditions in the NL group.

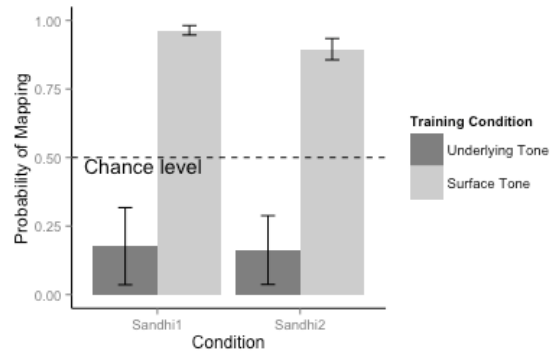


Figure 4: Probability of mapping across conditions in the BJ group.

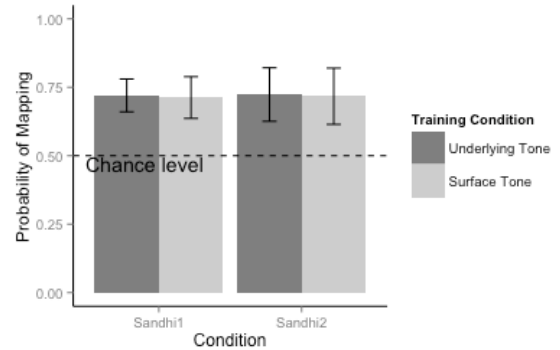


Figure 5: Probability of mapping across conditions in the NJ group.

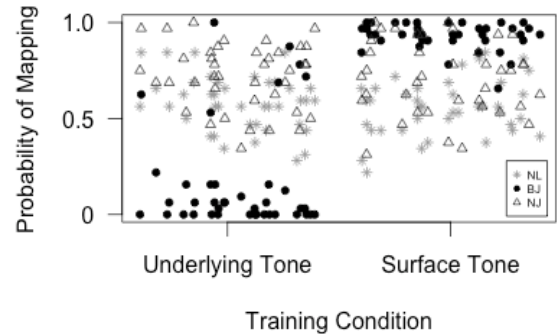


Figure 6: Probability of mapping for the three language groups across training conditions with Sandhi 1 and Sandhi 2 combined.

4. Discussion

4.1. NL group

Results from the NL group always fluctuated around chance level of 50% (see Figure 3). Also, this group did not demonstrate a better mapping in the surface-tone condition, where the training tone and test tone were acoustically nearly identical rendering the task less complex, as compared to the underlying tone condition ($p > .05$). This suggests that the NL listeners probably did not apply the newly-learned tone category in the test phase or worse, failed to successfully acquire the target tone category. For non-tone language listeners, it seems to be too difficult to generalize phonological representations for tones by purely acoustic exemplar-based listening, when faced with varying segmental features across exemplars. Possibly, the distraction from the tone on the second syllable also made phonologization of the tones on the first syllable challenging for them. Another reasonable explanation is that their short-term acoustical memories for tones were erased again during the time elapsed between the training and test when reading instructions.

NL group did slightly, but not significantly better in their mappings for Sandhi 1. This can be interpreted as an effect of phonological uninterpretability for tones in non-tone-language listeners outweighing any other possible effect of factors that were manipulated in this task, including the contrast between coarticulation and non-coarticulation.

4.2. BJ group

Distinct from the NL group, the results from the BJ group demonstrated a remarkable difference between the two training conditions ($p < .001$) (see Figure 4). A ceiling effect was observed in the mapping between surface tone and sandhied tone. Conversely, in the underlying-tone condition, most participants failed to accomplish the mapping although a few participants managed to do so, as Figure 6 clearly illustrates.

The performance of the BJ group can be explained by their native tonal grammar. The BJ group is known to have robust tonal categories [11, 12]. Four lexical tones occur in Beijing Mandarin, namely a high-level T1, a low-rising T2, a low-dipping T3 and a high-falling tone T4. As mentioned earlier, there is a T3 sandhi rule in Beijing Mandarin. On the one hand, the robust phonological representations for tones equip the BJ group to interpret the surface level of Nanjing tones in terms of their native Beijing tonal grammar remarkably well. On the other hand, this group has no access to the underlying level of Nanjing sandhied tones, because Sandhi 1 and Sandhi 2 in Nanjing are new phonological rules to them, and they lack the necessary mapping that would allow them to construct the surface pattern of a Nanjing sandhied tone based on its underlying properties. As a consequence, the BJ group always used a surface listening strategy in processing Nanjing sandhied tones, even though in the underlying-tone condition they were invited to use the underlying strategy.

No significant effect of coarticulation was found in the BJ group, possibly because this group focused exclusively on the tone on the first syllable while neglecting the tone on the second syllable as a context in the perception task.

4.3. NJ group

Results from the NJ group clustered around 75% mapping across all conditions. No significant difference was observed between the two training conditions ($p > .05$) (see Figure 5).

The NJ group is the only group in this study assumed to be equipped with both (underlying-based and surface-based) listening strategies. Hence, two competing listening strategies are at play in this group. The underlying-based strategy allows NJ listeners to successfully listen through the surface tone to access the underlying pattern in the underlying-tone condition by consulting the native grammar for NJ tone sandhi rules. This listening strategy also applies in the analysis of tones in the surface-tone condition, causing NJ listeners to still listen to sandhied tones at the underlying level and over-phonologize them. The surface-based strategy allowed NJ listeners to successfully match the phonetic realization of sandhied tones with surface tones in the surface-tone condition. This strategy also takes effect when incorrectly mapping the sandhied tone to its surface value at the acoustic level in the underlying-tone condition.

The results suggest that most of the NJ participants did not make a consistent choice from one of the two available strategies. Instead, they seem to be consistently mixing the two strategies in this perceptual task; whichever strategy they adopted subconsciously, the other strategy was always distracting them. Under continuous pressure from the two competing and interfering listening strategies, the NJ group ended up with a mapping significantly better than chance level compared to the NL group, but also remarkably poorer compared to the BJ listeners. An item analysis suggested no specific items caused NJ participants to fail in 100% mapping. Very few listeners consistently used a single listening strategy in doing the task, while for most NJ listeners the two listening strategies were at work competitively at the same time, although a preference for one strategy over the other was sometimes found.

In agreement with our prediction, no significant difference between the sandhi conditions was observed in the NJ group, because the native phonology always overrides any possible effect of coarticulation/non-coarticulation in this group.

5. Conclusions

In conclusion, results from the current study reveal distinct listening patterns in NL, BJ and NJ language groups. NL and BJ listeners listen to NJ sandhied tones at the surface level, with the BJ group interpreting NJ tones in terms of BJ tones highly successfully whereas the NL group experiences extreme difficulty in phonologizing NJ tones; NJ listeners may be equipped with both underlying and surface listening strategies and they seem to consistently mix the two strategies when perceiving sandhied tones. Apparently, the coarticulation property of the sandhi rule fails to bridge the underlying and surface levels for all groups of listeners in the current experiment. We suggest that this is due to the nature of the task employed here, which may have been too cognitively challenging in the case of the most promising NL listeners. Follow-up studies will explore the role of coarticulatory vs. non-coarticulatory tonal alternations in a different paradigm, aiming to make the task easier for NL listeners. Response latencies will also be analyzed, as a supplement to the current data.

6. Acknowledgements

Many thanks go to all the participants that participated in this experiment. We thank Dr. Ao Chen and Dr. Willemijn Heeren who give valuable feedback to this study. We also thank the support from the China Scholarship Council to the first author.

7. References

- [1] Huang, T., and Johnson, K.: 'Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners', *Phonetica*, 2010, 67, (4), pp. 243-267
- [2] Peng, S.-H.: 'Lexical versus 'phonological'representations of Mandarin Sandhi tones', *Papers in laboratory phonology V: Acquisition and the lexicon*, 2000, pp. 152-167
- [3] Francis, A.L., Ciocca, V., Ma, L., and Fenn, K.: 'Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers', *Journal of Phonetics*, 2008, 36, (2), pp. 268-294
- [4] Gandour, J.: 'Tone perception in far eastern-languages', *Journal of Phonetics*, 1983, 11, (2), pp. 149-175
- [5] Hallé, P.A., Chang, Y.-C., and Best, C.T.: 'Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners', *Journal of Phonetics*, 2004, 32, (3), pp. 395-421
- [6] Darcy, I., Ramus, F., Christophe, A., Kinzler, K., and Dupoux, E.: 'Phonological knowledge in compensation for native and non-native assimilation', *Variation and gradience in phonetics and phonology*, 2009, 14
- [7] Liu, D.: 'Nanjing Fangyan Cidian [Dictionary of the Nanjing Dialect]' (Jiangsu Education Publishing House, 1995)
- [8] Song, Y.: 'Nanjing Fangyan Shengdiao Shiyan Yanjiu (Experimental Study on Tones of Nanjing Dialect'. Master's dissertation, Nanjing Normal University, 2006
- [9] Jaeger, J.J.: 'Concept formation as a tool for linguistic research', *Experimental phonology*, 1986, pp. 211-237
- [10] Boersma, P., Weenink, D. : 'Praat: Doing phonetics by computer', 2010
- [11] Wang, W.S.-Y.: 'Language change', *Origins and evolution of language and speech (Annals of the New York Academy of Sciences)*, 1976, 280, pp. 61-72
- [12] Xu, Y., Gandour, J.T., and Francis, A.L.: 'Effects of language experience and stimulus complexity on the categorical perception of pitch direction', *The Journal of the Acoustical Society of America*, 2006, 120, (2), pp. 1063-1074