Language specificity and generality in the mapping between surface and underlying tones

Published by
LOT phone: +31 30 253 6111
Trans 10
3512 JK Utrecht e-mail: lot@uu.nl
The Netherlands http://www.lotschool.nl

Cover illustration: mapping between “surface” and “underlying” shapes of a tree and a boat, with falling leaves and changing clouds representing noises in surface forms, designed by Xin Li

ISBN: 978-94-6093-312-7
NUR 616

Language specificity and generality in the mapping between surface and underlying tones

Taalspecificiteit en -onafhankelijkheid in de mapping van lexicale tonen tussen onderliggende en oppervlaktevormen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

vrijdag 18 januari 2019 des middags te 2.30 uur

door

Xin Li

geboren op 3 februari 1989 te Anhui, China

Promotor: Prof. dr. R.W.J. Kager

This thesis was partly accomplished with financial support from the China Scholarship Council (CSC)

To my grandma

Table of Contents

Acknowledgement

At the end of this PhD journey, my heart is filled with joy and happiness, because of the knowledge and insights I gained in the past 5 years, and more importantly, because of those who I met and who supported me in the past 5 years.

René Kager is definitely on top of the list that I want to give thanks to. Without him this PhD journey would not have started. The luckiest thing for me in the past 5 years was to have him as my supervisor, and I can write a book about how much support I received from him. He was always available when I needed any kind of help, and he responded incredibly fast to all my emails. I was always looking forward to meetings with him, and my happiest time in a week was often after our meeting, when I left his office with loads of inspirations and motivations. He is the one who taught me how to be an independent researcher in every aspect. His enthusiasm for research, dedication to work, enormous patience with and respect to students, and kindness to all have deeply touched me and will forever guide me in my future life and career. René, you are my role model. I am so proud of being your student, and I hope to become someone like you.

I would like to express my sincere thanks to my advisors Willemijn Heeren, Ao Chen and Wentao Gu. Willemijn, thank you for being there whenever I asked for help. I learned a lot from all the suggestions you gave on my experiments and my manuscript. Ao, thank you for the insightful discussions on experiment settings and thought-provoking questions at an early stage of my PhD project. Wentao, thank you for providing me with your phonetic lab in Nanjing Normal University, for your suggestions on my Chapter 3, and for your general interest in what I am doing.

I hope to thank many professors in and outside UiL-OTS who either allowed me to attend their courses or helped me in various ways. From them I learn how attractive a lecture can be and how generous they are in helping young researchers. I respect them from the bottom of my heart. Prof. Hugo Quené, thank you for your introduction of R to me in your statistics course, and also for your kind replies to my emails about growth curve analysis. Prof. Huub van den Bergh, thank you for the private meeting with me on statistics analysis and for introducing MLwiN to me. Prof. Vincent van Heaven, thank you for the prosody course in 2014. Though I didn't grab an opportunity to speak to you after class, I sincerely admire how devoted you were to the class, both in your classroom and in your long emails before and after class. Prof. Carlos Gussenhoven, thank you also for this prosody course, for the lovely dinner at your house, and for your questions, supports and encouragements at various international conferences. Meeting you every time was truly a great pleasure. Prof. Aditi Lahiri, thank you for the private session with me at the TIE 2016 conference after you missed my talk, and for your suggestions on my future studies.

My sincere gratitude also goes to my reading committee members. They are Prof. Patrick Wong, Dr. Yiya Chen, Prof. Sharon Peppercamp, Prof. Janet Grijzenhout, and Prof. Aoju Chen. Thank you for spending time on reading this dissertation and for giving valuable comments.

I owe a special *thanks* to all the technical and methodological support people in UiL-OTS. Chris (van run), I worked with you on almost all the experiments in my project. I was amazed by how efficient you were at solving problems, how much you thought in advance for the experimenter, and how kind you were when replying to emails. You were so reliable and so nice to work with. Maarten (Duijndam), thank you also for your timely help on fixing experiment bugs and faulty equipments. Iris (Mulders), thank you for sending invitations for all my experiments. Kirsten (Schutter), thank you for the countless discussions on statistical issues via emails and skype & office meetings. Thank you for your expertise, your always willingness to help, and your patience.

Many thanks also go to the administrative people in UiL-OTS. They include at least our research coordinator Maaike Schoorlemmer, our institute director Frank

Wijnen, our secretaries and janitors. Maaike, I still remember the first time I met you in Beijing in the winter of 2012. You are the first person I know from UiL-OTS. Thank you for the hug you gave me when you heard I had received the scholarship, and for the countless help you provided in the past years. Frank, I was impressed by how responsible you are as our institute director and how grateful you were for people's work in the site visit of 2018. Yvonne van Adrichem, Mariélle Hilkens, and Zarah Farhan, thank you for your help here and there. Barend (van Manen), thank you for your politeness and your generosity. Merryl (van Zichem), thank you for your big smiles and greetings on many late evenings when I was probably the only one in Trans, which made me feel less lonely. There are also other administrative people in Utrecht University that deserve a *thank you*. I always received quick, kind and helpful replies wherever I wrote to.

I was probably the person who spent most time on the train (between Utrecht and Groningen) in the Netherlands from 2014 to 2017. The 4 hours I spent on the train per day allowed me plenty of time to do self-reflection. Yet it prevented me to spend more time with my friends and colleagues in UiL-OTS. I have always wished to host you in Groningen, but I almost never had such opportunities. I hope to thank all those who accept me as a friend in spite of so little time I spent with you. Carolien (van den Hazelkamp), my first office mate, thank you for approaching me when I was shy, for sharing family stories, and for being so helpful whenever I had questions. Anne (van Leeuwen), thank you for being such a good model of PhD student and for being so easygoing in the office. Hanna (de Vries), thank you for the lovely gift and message you left on my desk. They were very heart-warming. Bei (Peng), I was so excited when I had you beside me in my office, and I spent wonderful time with you during your short stay. Bambang (Kartono), thank you for hours of talk everyday and for your always willingness to accompany me. Merel (Scholman), thank you for showing me the bakery which soon became my favourite, and for kindly sharing with me tips for defense preparation. Corentin (Bourdeau), it was always a great pleasure to see you in office, to talk with you, and to see you dance. I wish we had been office mates earlier. Bambang and Corentin, the drinks and dinners we had together after work made the last few months of my PhD life most unforgettable. You are amazing friends. Thank you my dear office mates above for all the time we spent together in our isolated office 2.34. Zenghui (Liu), thank you for providing me with a bed several times, for showing me how kind a person can be, and for your trust in me to be your paranymph. You are like a sister to me. Mengru (Han), thank you for your encouragements and suggestions in the last months before I submitted my dissertation, for the chats we had late at night, and for sharing photos of Ally. Shuangshuang (Hu), thank you for sharing cute pictures and videos now and again. Life would have been boring and unbearable without you. Xiaoli (Dong) and Bert (Le Bruyn), thank you for the lovely dinners at your house. You are the best hosts. Xiaoli, thank you for always being so nice to me. Jingwei (Zhang), thank you for kindly offering me your lab to do experiment in Nanjing University, and for your help in recruiting participants. Luying (Hou), though we didn't meet as often as we wanted, I know how kind you are and I always feel close to you. Yipu (Wei), thank you for showing me around in Trans on my first day here and made me feel included. Fang (Li), thank you for advising me at the beginning and at the end of my PhD. Anqi (Yang), thank you for voluntarily helping me with Praat scripts. Shanshan (Fan), thank you for your company when I visited Beijing. Liquan (Liu), thank you for your suggestions on my experiment and for your humor. Marlisa (Hommel), thank you for visiting my place in Groningen, for helping me with the Dutch summary of this book when I felt desperate, and for being the best guide in Amsterdam. Andrea (Santana Covarrubias) and Stavroula (Alexandropoulou), thank you for the wonderful time we had in organizing ELiTU. Myrthe (Bergstra), thank you for warming my heart with kind emails and words. Björn ('t Hart), thank you for being gentle, kind and helpful. Thanks also to all the other friendly faces around Trans in the past years, including Na Hu, Chou Mo, Yingtong Liu, Yuan Tian, Emily Ge, Maartje Schulpen, Jolien Scholten, Quy Ngoc Thi Doan, Maartje de Klerk, Heidi Klockmann, Brigitta Keij, Anne-France Pinget, Martijn van der Klis, Emma Everaert, Iris Selten, Leslie Piggott, Anja Goldschmidt, Nina Sangers, Suzanne Kleijn, Franca Wesseling, and many more.

My paranymphs Yuan (Xie) and Jeroen (Breteler), thank you both for agreeing to accompany me in my defense without any hesitation. You are very trustworthy friends to me. Yuan, there is always a magic wind blowing me to your office to say hi. You are always ready to help, which touched me deeply. Jeroen, thank you for your frequent visits to my office, for sharing photos and thoughts, for your generous help in translating experiment instructions for my Dutch participants, and for always being so considerate.

I also thank my friends and colleagues from other linguistic institutes. They are Qian Li, Ting Zou, Yang Yang, Min Liu, Yifei Bi, Junru Wu, Menghui Shi, Ziyi Bai, Han Hu and Qing Yang from Leiden University, Lei Wang from Radboud University Nijmegen, Yan Xia and Yueqiao Han from Tilburg University, and Beinan Zhou, Cong Zhang, Xiaole Sun, Yang Li and Mengyue Wu, whom I met at conferences. We witnessed the progress of each other in the past years and hope we will always keep in touch. Among them, I thank Lei Wang in particular for the insightful discussion on data analysis, and Yan Xia in particular for being the speaker for my experiment.

Jun Zhang and Xiaorong Lu also deserve a special *thank you*, for their kind help with recruiting participants for my experiments in Nanjing and Bejing, respectively. My experiments would not have completed without their help.

Finally, I thank my family for forever being my strongest support. My dad Yin Li and my mom Shunsheng Hu, thank you for giving me life and for always believing in me. Your two visits to the Netherlands and our travel around Europe are my happiest memories in the last 5 years. I am so glad that I can stay closer to you again soon. My parents in law Meilin Lu and Yunlong Ma, you have treated me as a daughter all the time. I cannot be luckier to have you as family. Shenghui Wang (my sister-like auntie) and Gwenn Englebienne, thank you for hosting us in Amsterdam. I was happy to spend two Spring Festivals with you and your family, and to witness the growth of your beautiful children. Mushroom my chinchilla, Pangpang, Pompom and Keji my rabbits, you filled my life with comfort and joy. I could not imagine my life without you little cuties. Chao my husband, thank you for taking care of me in life, for listening and for sharing. Without you I would never have the courage to begin my PhD in this country, and I would not have completed this dissertation. We may be in two different countries in the coming 1 or 2 years. However, love and faith will keep us connected. Let’s fight together for a brighter future.

My grandma Aizhi Sun, I think of you more and more often as I grow older. How I wish you were still with me, and I could take care of you and show you around this beautiful country. This book is a late gift for you.

Chapter 1

Introduction

1.1 Dissertation goals

In connected speech, the production of a speech sound is often affected by the articulatory features of its neighboring sounds (e.g., Beddor & Krakow, 1999; Fowler, 1980; Hammarberg, 1976; Lahiri & Reetz, 2002; Xu, 1994); this typically results in a different *surface* (phonetic) realization of the sound from its *underlying* form, for example, in Korean, an underlying labial may surface as a velar when immediately preceding another velar (Kim-Renaud, 1974).

Mismatches between surface sound and underlying sound due to contextual change in production may pose a problem in perception for listeners (Beddor, 2009). How do listeners perceive surface sounds? Are they able to map the surface sounds to the underlying sounds? The current dissertation sets out to investigate the specificity and generality of the knowledge that allows listeners to perform surface-to-underlying sound mapping. In order to approach this research goal, I ask the following research questions: 1) Is the surface-to-underlying mapping dependent exclusively on language-specific knowledge, and is it therefore restricted to native listeners? Or is this mapping also assisted by language-general articulatory knowledge, and for this reason, also accessible to naïve non-native listeners? 2) Under what circumstances is the surface-to-underlying mapping accessible to naïve non-native listeners? Is the difficulty of the surface-to-underlying mapping influenced by different types of processes, e.g., assimilatory versus dissimilatory processes? Is the mapping by naïve non-native listeners also affected by the categoricalness versus gradience of assimilatory and dissimilatory processes?

Native listeners and naïve non-native listeners are usually hypothesized to perform the mapping from the surface sound to the underlying sound in completely different ways. Taking the Korean labial-to-velar assimilation rule (e.g., /ip+ko/ ‘wear and…’ becomes [ikko] (Jun, 2004)) as a hypothetical case, the surface form, a velar-velar sequence, contains information that is ambiguous only for native listeners (and not ambiguous for non-natives), because underlyingly it can be either (a) a labial-velar sequence, in which the labial is neutralized to a velar under the influence of the contextual (following) velar, or (b) a true velar-velar sequence. For native listeners, to perform surface-to-underlying mapping means to ‘undo’ the neutralizing labial-to-velar neutralization process and recover the underlying identity from the surface form. According to the standard view, native listeners have accumulated native lexical knowledge or knowledge of the phonological process from years of experience, which allows them to undo the neutralization. In contrast, for naïve non-native listeners, this language-specific phonological knowledge or lexical underlying representations is unavailable. According to an alternative view, which will be investigated in this dissertation, what naïve non-native listeners are hypothetically doing in a surface-to-underlying mapping task, if they are able to do it, is construing a link between the surface sound and the target (underlying) sound via the given context, presumably through some language-universal articulatory knowledge.

The specificity versus generality of the knowledge that supports surface-to-underlying mapping can be tested through assimilatory and dissimilatory processes. The influences imposed by the contextual sounds can be *assimilatory*, when the underlying sounds receive extended acoustic features from the contextual sounds, and *dissimilatory*, when the added acoustic features on the underlying sounds conflict with the contextual sounds. Assimilatory processes are mostly thought to be grounded in articulatory ease (e.g., Grammont, 1933; Johnson, 1973). Listeners’ implicit knowledge about articulatory gestures, either from native articulatory experience, or from general articulatory settings, is, to some extent at least, language general. As a result, the surface-to-underlying sound mapping for assimilatory processes is more likely to be language-general (the "language-general mapping hypothesis for assimilatory processes"). In contrast, dissimilatory processes are mostly thought to originate either from listeners' "hyper-correction" (Ohala, 1993), an account which implies listeners' knowledge of language-specific underlying representations, or alternatively, from speakers' difficulties in the "motor planning" of sequences of identical or similar sounds (Frisch, 2004; Frisch, Pierrehumbert, & Broe, 2004; Garrett, 2015; Garrett & Johnson, 2012; Grammont, 1895, 1933; Tilsen, 2007), an account which is less clear in the language specifity/generality of liseners' knowledge it implies. Since it turns out to be rather difficult to derive testable predictions from the "motor planning" view, I will only test the hypothesis clearly predicted by the "hyper-correction" view, i.e., the surface-to-underlying mapping needed for dissimilatory processes is likely to require native experience of the language (the "language-specific mapping hypothesis for dissimilatory processes"). (Detailed discussion on assimilatory and dissimilatory processes will be given in Section 1.2.)

As a specification of the language-general mapping hypothesis for assimilatory processes, a surface-to-underlying mapping may be available to non-native listeners for an assimilatory process only in case the underlying representation is to some extent acoustically recoverable from the surface (the "gradient-based language-general mapping hypothesis for assimilatory processes"). That is, the influences imposed by the contextual sounds can be either *gradient*, leaving some acoustic traces of the underlying element, or *categorical*, leading to a phonemic change, often a neutralization. (Details on the gradient/categorical nature of changes will be found in Section 1.2.) The *categoricalness* versus *gradience* of a process may also influence whether the mapping between the surface and the underlying sounds is accessible for different types of listeners. For native listeners, a categorical change should pose no absolute obstacle for phonologically recovering the underlying element; whereas for naïve non-native listeners, the presence of acoustic residues of the underlying sound should be crucial to establishing any mapping between the surface and underlying forms.

This dissertation will look into lexical tone and tone sandhi phenomena to investigate the above issues. (Details on tone and tone sandhi will occur in Section

1.3.) In order to subject the language-general mapping hypothesis for assimilatory processes to the strongest possible test, I intend to keep the non-native listener group as naïve as possible. By using the phenomena of lexical tone and tone sandhi, I will be able to test a group of naïve non-native listeners that is maximally distant from the native language group, namely a non-tonal language group without any previous exposure to tones, hence guaranteed to have no experience with tone or tone sandhi. From many studies, it is known that non-tonal-language listeners have severe problems in perceiving lexical tones, e.g., Francis, Ciocca, Ma, and Fenn (2008) found that naïve English listeners identified Cantonese tones correctly only 66% of the time.

To summarize, the main goal of this dissertation is to investigate the specificity versus generality of listeners' mapping between surface sounds and underlying sounds. Specifically, the mapping problem will be examined along three dimensions: 1) whether the mapping depends on language-specific knowledge, or whether it is facilitated by language-general articulatory knowledge as well; 2) whether the difficulty of mapping differs for assimilatory and dissimilatory processes, especially for naïve non-native listeners; 3) as a refinement of 2), whether the mapping is further influenced by categorical and gradient processes, for naïve non-native listeners. This dissertation will examine the three aspects of the mapping issue in the phenomena of tone and tone sandhi.

The rest of this introductory chapter will be mainly devoted to reviewing literatures that are relevant to the above goals. First, in Section 1.2, I will review studies on assimilatory/dissimilatory and categorical/gradient processes from a production perspective. In Section 1.3, I will introduce tone and tone sandhi phenomena, including assimilatory/dissimilatory, and categorical/gradient cases, again from a production angle. Next, Sections 1.4 and 1.5 will review segmental and tonal studies on listeners' surface-to-underlying mapping for assimilatory/dissimilatory and categorical/gradient processes, with Section 1.4 focusing on native mapping, and Section 1.5 on naïve non-native mapping. Finally, after reviewing these literatures, in Section 1.6, I will present the research questions, explain how these will be addressed by individual studies, and propose the outline of this dissertation.

1.2 Assimilatory and dissimilatory processes

The production of one sound in natural continuous speech is often affected by the articulatory features of its neighboring sounds. When this contextual change happens, the articulated surface form of the target sound differs from its underlying form. For instance, an underlying oral vowel may surface as nasalized due to coarticulation with an adjacent nasal consonant, as in Thai and English; an underlying coronal may surface with labial place when immediately followed by a labial, as in English, *lean bacon* can be heard as *lea*[m] *bacon* (Gaskell & Marslen-Wilson, 1996); or a labial may become a velar when immediately preceding a velar, as in Korean, /ip+ko/ 'wear and…' can surface as [ikko] (Jun, 2004) (Kim-Renaud, 1974).

For purposes of this dissertation, the influences imposed by the contextual sounds can be grouped by two dimensions: gradient/categorical and assimilatory/dissimilatory. A contextual process can result in a *categorical* change, in which case the articulatory gesture or characteristic of the underlying element is completely lost in the production of the surface form, and the derived element has become neutralized with another sound. For example, in the Korean labial-to-velar change, the surface labial resembles an underlying velar with extensive gestural overlap (Jun, 2004). A process can also result in a *gradient* change, in which case some articulatory gesture of the underlying sound remains in the surface form. For example, in English coronal place assimilation, coarticulated coronals were found to have traces of the coronal segments in formant features (Gow, 2003).

The processes resulting in contextual changes can be classified as *assimilatory* or *dissimilatory* based on whether the altered sound becomes more or less similar to its context sound. When the context sound extends its articulatory features to the coarticulated sound, the process is defined as *assimilatory*. All the examples I listed earlier in this section are assimilatory in nature, e.g., the nasalization of a vowel immediately before a nasal context in languages such as English and Thai. When the context sound causes the target sound to change its phonetic features while these sounds underlyingly share the same phonetic features, the process is referred to as *dissimilatory*. An example is found in Southern Bavarian German, where the liquid

/r/ is converted to a non-liquid /d/ before or after an adjacent liquid /l/, e.g., /tiər-lə/ ‘animal (diminutive)’ becomes [tiədlə] (Hall, 2009). This dissertation is mainly concerned with assimilatory and dissimilatory processes that are locally conditioned, i.e., changes triggered by an adjacent sound, although non-local contextual assimilatory and dissimilatory changes are also observed in natural languages, e.g., non-local liquid dissimilation in Latin (e.g., /sol-alis/ ‘solar’ becomes [solaris]) (Abrego-Collier, 2013)), and non-local liquid assimilation in the Bantu language Bukusu (e.g., /kar-il-a/ ‘twist’ becomes [kar-ir-a] (Odden, 1994)).

While assimilatory and dissimilatory processes are well attested in many languages, assimilatory phenomena are far more prevalent than dissimilatory phenomena across the languages of the world (e.g., Cutler, 1998; Johnson, 1973). This difference in their prevalence may be partly due to their origins.

Assimilatory processes are mostly believed to be grounded in articulation. Many scholars have conceived these as motivated by increasing ease of articulation (e.g., Grammont, 1933; Johnson, 1973). Among the various views accounting for dissimilatory changes, the most widely accepted view was proposed by Ohala (1993), who posited that a dissimilatory sound change occurs because listeners misattribute a feature that is intrinsic to the surface form to coarticulation, erroneously applying corrective processes (“hyper-correction”). The “hyper-correction” view holds that listeners erroneously recover a speaker’s intended pronunciation, misattributing it to an underlying sound which (by definition) instantiates an existing phonological category in the language. An alternative classic view on the motivation of dissimilatory processes is the “motor planning” account (Frisch, 2004; Frisch et al., 2004; Garrett, 2015; Garrett & Johnson, 2012; Grammont, 1895, 1933; Tilsen, 2007). “Motor planning” is generally viewed as the process of “constructing or retrieving motor plans that will later be executed by speaking” (Garrett & Johnson, 2012, p. 59). According to Garrett and Johnson (2012), this view was first proposed by Grammont (Grammont, 1895, 1933), who argued (as cited in Garrett and Johnson (2012, p. 57)) that non-local dissimilation occurs “when planning for a segment in a more prominent position distracts a talker who is producing a similar segment in a weaker position.” For example, Frisch (2004, p. 346) and Frisch et al. (2004, p. 180), proposed that dissimilation may occur because of the difficulty of processing identical or similar sounds, a scenario interpreted by Garrett and Johnson (2012) as also belonging to "motor planning." Perhaps the most insightful discussion of "motor planning" as the motivation for dissimilation was offered by Garrett (Garrett, 2015; Garrett & Johnson, 2012). Garrett and Johnson (2012) proposed that (non-local) dissimilation is caused by the speaker's "inhibition of one segment by the activation of another" due to the preference of an alternating sound pattern over a repetitive pattern (p. 60).

From the different origins of assimilatory and dissimilatory processes, the language-general/specific nature of surface-to-underlying mapping in these processes can be hypothesized. The Motor Theory (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985) and the Direct Realist Theory (Fowler, 1986, 1996, 2006; Fowler & Brown, 2000) agree on the idea that acoustic events are perceived as being caused by articulatory gestures. Evidence can be found compatible with this idea. For instance, McGurk and MacDonald (1976) found that the perception of a speech sound is distracted by the visual information of a mouth performing the articulatory gestures of another sound (e.g., /ba/ is perceived as /da/ when a mouth articulating /ga/ is presented) (the "McGurk effect"). Moreover, D'Ausilio et al. (2009) found that magnetic stimulation of lip-related areas in the motor cortex facilitated discrimination of lip-related speech sounds (/ba/, /pa/) but not tongue-related sounds (/da/, /ta/) presented in noise, whereas stimulation of tongue-related motor cortex areas produced the reverse effect. The idea that articulatory ease provides the motivation for assimilatory processes would suggest that the surface-to-underlying mapping for these processes is possibly language-general, i.e., accessible for both native and naïve non-native listeners of the target language. The strongest test of this idea resides in processing by naïve non-native listeners. It is hypothesized here that what these listeners might be plausibly doing when mapping assimilatory processes is to map a surface form onto a hypothetical underlying sound, presumably by means of implicit knowledge of articulatory gestures. Such knowledge may be generalized from experience with their native language, or alternatively, may even exist independently of native language experience, e.g., by relating non-native sounds to articulatory settings by means of some form of introspection, without having experience with the articulation (the language-general mapping hypothesis for assimilatory processes).

Regarding dissimilatory processes, the "hyper-correction" view (Ohala, 1993) holds that listeners erroneously recover a speaker's intended pronunciation, misattributing it to an underlying sound which (by definition) instantiates an existing phonological category in the language (as argued above). This view implies that the surface-to-underlying sound mapping for dissimilatory processes should be informed by language-specific knowledge, i.e., that this mapping should be only accessible for native listeners, and inaccessible for naïve non-native listeners, because the lexical underlying representations that this view crucially refers to are, by definition, unavailable for naïve non-native listeners (the language-specific mapping hypothesis for dissimilatory processes). Alternatively, considering the "motor planning" approach (Frisch, 2004; Frisch et al., 2004; Garrett, 2015; Garrett & Johnson, 2012; Grammont, 1895, 1933; Tilsen, 2007) to dissimilatory processes, two tentative predictions might be developed. On the one hand, if knowledge of production plays a role in perception (mapping), then speakers' general non-language-specific dispreference for similar/identical sound sequences in production may help the speaker-turned-into-listener to successfully undo dissimilatory processes; on the other hand, surface dissimilar sound sequences may massively overcue naïve listeners, making natural speech difficult to process, because natural speech is full of dissimilar sounds which do not originate from dissimilation. Taking together, the "motor planning" view does not derive consistent and clear predictions regarding the language specificity/generality of the mapping. Consequently, in this dissertation, I will only test the language-specific mapping hypothesis undisputedly predicted by the "hyper-correction" view of dissimilatory processes.

Currently, experimental evidence supporting the language-general surface-to-underlying mapping hypothesis for assimilatory processes and the language-specific mapping hypothesis for dissimilatory processes is missing. A comparison between the surface-to-underlying mapping for native and non-native listeners may reveal the generality versus language experience in terms of mapping in assimilatory and dissimilatory processes.

As discussed in Section 1.1, this dissertation will investigate these hypotheses for tone and tone sandhi phenomena, to achieve maximal naïveness of the non-native group. The next section will introduce tone and tone sandhi phenomena, including assimilatory/dissimilatory cases and categorical/gradient cases, from a production perspective.

1.3 Tone and tone sandhi

1.3.1 Tone

Tones can be viewed as phonemic distinctions at a suprasegmental level. They serve a lexically contrastive role in tonal languages such as Mandarin Chinese, Cantonese, Thai and Vietnamese. For example, in Mandarin Chinese, the syllable *ma* bearing four different tones means ‘mother,’ ‘hemp,’ ‘horse,’ and ‘to scold,’ respectively.

Tones are mainly realized by pitch (f0) variations. In terms of directionality, they can be static (level); or dynamic (contour), the latter including falling, rising, and complex contours which change direction in the middle, e.g., falling-rising. Tones with the same direction can be further classified by a) steepness, e.g., in a pair of rising tones sharing the same onsets but differing in offsets, the rising tone with the higher offset is steeper; and b) pitch register, e.g., a pair of falling tones with the same steepness may have different overall heights.

Tones can be indicated by numbers on Chao (1930)’s five-level scale, with the number 1 representing the lowest pitch level in the speaker’s pitch range, and 5 representing the highest. For example, Mandarin Chinese has four lexical tones: high level 55 (Tone 1), mid/high-rising 35 (Tone 2), low falling-rising/dipping 214 (Tone 3) and high falling 51 (Tone 4) (Chao, 1968).

In tonal languages such as Mandarin, Cantonese and Thai, lexical tones are associated with syllables (Xu, 1998). The domain of tone inside a syllable is widely debated and may vary between languages (e.g., Howie, 1974; M. Lin & Yan, 1995). Most recent researchers of Chinese tonal languages adopt the rhyme of the syllable, including the syllable nucleus and coda, as the domain of tone (e.g., Gu, 2015; Zeng & Mattys, 2016; Jingwei Zhang, 2014).

1.3.2 Tone sandhi

Just like consonants and vowels, tones can also undergo variations induced by their tonal contexts in connected speech. When this tonal contextual change happens, the tone's canonical f0 contour can be modified, leading to altered phonetic realizations at surface level. Tonal contextual changes are widely found in many tonal languages, such as Beijing Mandarin (Cheng, 1973; Yip, 1980, 2002), Tianjin Mandarin (Q. Li & Chen, 2016; Jie Zhang & Liu, 2011), Taiwan Southern Min (Myers & Tsay, 2008), Thai (Abramson, 1979), Malaysian Hokkien (Chang & Hsieh, 2012) and Vietnamese (Han & Kim, 1974).

When the modification induced by the tonal context renders a lexical tone to change its tonal category, and the derived tone is perceptually indistinguishable from another lexical tone, it is usually referred to as "tone sandhi" (Xu, 1994). In this dissertation, I will use "tone sandhi" in a broader way for convenience to include both *categorical* and *gradient* tonal changes induced by the context. As an example of a categorical sandhi process, Taiwan Southern Min has a tone sandhi rule which changes a high level *yinping* (in traditional tone classification of this dialect) preceding a low rising *yangping* categorically to a mid level *yangqu* (Myers & Tsay, 2008). This tone sandhi process can be written as 55.24 → 33.24 including tone values (the dot is used to separate pitch values of the tones). A gradient tone sandhi rule example is seen in Tianjin Mandarin, in which a high-falling T4 preceding a mid-falling T1 turns into a high-rising tone which resembles a lexical mid-rising T2 but keeps its original high onset (Q. Li & Chen, 2016).

For purposes of this dissertation, tone sandhi processes can also be classified as *assimilatory* and *dissimilatory*. For example, in Zhenjiang Mandarin (Qiu, 2012), an assimilatory tone sandhi rule turns a high falling Tone 1 adjacently before another high falling Tone 1 into a high level tone, with its tone offset raised to the same pitch height as the following tone's onset (42.42 → 44.42). Beijing Mandarin has a Tone 3 sandhi rule, which is an example of the dissimilatory tone sandhi process (Cheng, 1973; Yip, 1980, 2002): a low dipping Tone 3 immediately preceding another low dipping Tone 3 changes to a low-rising Tone 2 (214.214 → 24.214). The distinctive feature of lexical Mandarin Tone 3 is "low," and this tone sandhi rule is often transcribed as LL.LL → LH.LL (e.g., Yip, 1980), with the altered tone offset deviating from the following tone's onset.

1.4 Surface-to-underlying mapping by native listeners

Several experimental studies have shown that upon hearing surface forms in speech, native listeners are able to map a surface sound to its underlying counterpart by perceptually undoing the neutralization effects that can be attributed to context and successfully recovering the underlying sound. For instance, English native listeners were found to perceptually attribute coarticulatory vowel nasalization to the immediately following nasal consonant context and recover the underlying vowel. For example, Lahiri and Marslen-Wilson (1991) in a gating study found that when presented with a surface nasal vowel, native English listeners anticipated the upcoming nasal consonant, which suggested that they attributed the nasality on the vowel to the nasal context and interpreted the nasalized vowel as an underlying oral vowel; Beddor and Krakow (1999) found that native English listeners showed difficulties in judging the nasality of nasal vowels embedded in nasal context, and often identified the nasal vowels as underlying oral vowels.

1.4.1 Native mapping in assimilatory processes

Native listeners are found to be able to map the surface sounds to their underlying identities in assimilatory processes in various languages across the world. In addition to the English vowel nasalization example mentioned above, more examples can be found. For instance, as a reversed version of the coronal place assimilation process in English, German and Dutch, native listeners undo the neutralization of place due to the context sound and perceive a word-final labial, when followed by a word-initial labial, as a coronal (German: Coenen, Zwitserlood, & Bölte, 2001; English: Darcy, Ramus, Christophe, Kinzler, & Dupoux, 2009; Gaskell & Marslen-Wilson, 1996, 1998; Gaskell & Snoeren, 2008; Gow, 2003; Dutch: Mitterer, 2003), e.g., *lean bacon* is successfully recovered from *lea*[m] *bacon* (Gaskell & Marslen-Wilson, 1996). Examples can also be seen in voicing assimilation in Hungarian (Gow & Im, 2004), French (Darcy et al., 2009), and

German (Coenen et al., 2001), and labial-to-velar assimilation in Korean (Mitterer, Kim, & Cho, 2013).

These studies involved not only gradient processes, when an articulatory gesture of the underlying form partially remains in the surface form, but also categorical processes, when an articulatory gesture of the underlying form is completely lost in the surface form. For example, the coronal place assimilation used in Gow (2003) and Mitterer and Blomert (2003) was a gradient process, whereas extreme assimilated forms of this assimilation were used in Gaskell and Marslen-Wilson (1996), Gaskell and Marslen-Wilson (1998), Coenen et al. (2001), Gaskell and Snoeren (2008) and Darcy et al. (2009). The Hungarian voicing assimilation used in Gow and Im (2004) was also a gradient process. However, the French voicing assimilation used in Darcy et al. (2009), the German voicing assimilation used in Coenen et al. (2001), and the Korean labial-to-velar assimilation used in Mitterer et al. (2013) all resulted in categorical phonemic changes. Among these studies, Mitterer et al. (2013) found that the most assimilated sounds, which were most ambiguous to native listeners, lead to more successful surface-to-underlying mappings for the listeners, compared to the less assimilated sounds. They created a six-step continuum between a natural Korean labial end point and a natural Korean velar end point, both taken from isolated forms. Each step on the continuum was presented either with a "*viable context*," which allows for the labial-to-velar change or an "*unviable context*," which does not license the change. Each step was categorized by native listeners as belonging to either the labial or velar category. The authors observed that the native listeners categorized the velar end of the continuum as labial in the viable context at a significantly higher rate than in the unviable context. Yet in the middle part of the continuum, where the test signal sounded intermediate between a labial and a velar, the viable context did not produce such a facilitating effect for the native listeners when mapping towards the underlying labial. This set of studies suggested that the native surface-to-underlying mapping in assimilatory processes probably does not rely on the presence of residual phonetic cues for the underlying sound in the surface pattern, but rather relies on the lexical knowledge or phonological knowledge acquired in one's native language.

Native listeners were observed to only perform the surface-to-underlying mapping in a context-sensitive way, i.e., when the context that licenses the underlying-to-surface change is present. This observation is the counterpart of what happens in contextual sound production, e.g., *lean* will only be pronounced as *lea*[m] when followed by a word starting with a labial, but not when followed by words whose initial consonants have another place of articulation. For instance, Gaskell and Marslen-Wilson (1996) found in a cross-modal priming task that when native English listeners heard altered tokens like /wɪkɪb/ embedded in contexts that permitted the change as a result of a place assimilation process (e.g., /wɪkɪb præŋk/) and then saw the visual target (e.g., *wicked*), the priming effect was very strong; whereas when they heard /wɪkɪb/ in neutral contexts like a velar (e.g., /wɪkɪb geɪm/), this did not effectively cause a priming effect. In another cross-modal priming study, Coenen et al. (2001) tested native German listeners on German place assimilation (e.g., *bring tulpen* 'to bring tulips' becomes *bring* [k]*ulpen*) and voice assimilation processes, and also found large priming effects toward the underlying lexical target in the appropriate context; on the contrary, no priming effect was observed when the changed words were presented in isolation. Mitterer et al. (2013) tested the Korean labial-to-velar assimilation process, and observed in an eye-tracking paradigm that native listeners gave more looks and faster reactions to the labial target when hearing the velar in contexts that allowed for the labial-to-velar assimilation than in neutral contexts. This context sensitivity suggests that the mapping between the surface and the underlying sounds should not be a purely lexical inference process, but rather, it involves phonological inference.

Some studies engineered experimental conditions in a certain way that lexical knowledge could not be made use of in the mapping task. For instance, Gaskell and Marslen-Wilson (1998) adopted non-lexical words such as *preight* in a phoneme monitoring experiment, and found that native English participants heard /t/ (a part of /preɪt/ (*preight*)) more often in /preɪp beərə/ (*preight bearer*), where the context supported the place assimilation, rather than in /preɪk beərə/, where there was no clue about the place assimilation. The results indicated that the phonological inference process which recovers the underlying form of speech operates also on non-words, suggesting that the native mapping is not based entirely on access to lexical knowledge.

Tonal studies added to the evidence that native listeners use the variation due to assimilatory processes in a phonological way to recover the underlying tone identity.

For example, Xu (1994) had native Mandarin speakers naturally produce trisyllabic words with rising (Tone 2) and falling (Tone 4) tones on the middle syllable, and then segmentally modified the first and third syllable to obscure the semantic status of the trisyllabic word (e.g., /tɕʰiT4 ɕiaŋT4 ʈʂanT4/ 'weather station' was modified to /tʰiT4 ɕiaŋT4 ʈʂanT4/ (nonsense string). He then presented the trisyllabic nonwords to native Beijing Mandarin listeners and asked them to identify the tonal identity of the changed rising/falling tone on the middle syllable. Although the rising/falling tone in the middle underwent gradient and categorical changes due to different tonal contexts, the native listeners recovered their underlying identity at an overall high accuracy, indicating that this surface-to-underlying mapping did not rely on phonetic residues. This study also included a condition when the context of the coarticulated tone was replaced by white noise, when it was observed that the identification accuracy for the more categorically changed tones dropped to below chance level, suggesting that the mapping was context dependent. Nonsense strings were used which would not remind listeners of any real trisyllabic words, thus providing evidence that abstract phonological knowledge was used in the mapping.

1.4.2 Dissimilatory processes

Regarding dissimilatory processes, few segmental studies have explored native listeners' capability of mapping the surface form to the underlying representations. Some tonal studies, such as Peng (2000), A. Chen and Kager (2011), and A. Chen, Liu, and Kager (2015) investigated the Mandarin Tone 3 sandhi rule (T3.T3 → T2.T3), which is considered dissimilatory in nature (e.g., Cheng, 1973; Yip, 1980, 2002) as discussed in Section 1.3, and provided some evidence in favor of the sandhied tone being perceived by native listeners as an underlying Tone 3.

Peng (2000) tested native Mandarin listeners on the Mandarin Tone 3 sandhi process in a Concept Formation experiment (this paradigm will be discussed in detail in Chapter 3). They first trained the listeners with the tonal category of Tone 3, and then presented them with the derived Tone 2 (sandhied Tone 3) of the Mandarin Tone 3 sandhi rule followed directly by the contextual Tone 3, and asked them to categorize the derived tone. The listeners immediately and consistently (> 80%; chance level: 50%) categorized the surface Tone 2 as belonging to its underlying category Tone 3. This study did not include a control context condition, and hence did not illuminate if the mapping from the surface tone to the underlying tone was context-sensitive. It adopted test words whose surface disyllabic words do not exist in the Mandarin lexicon whereas their underlying disyllabic words do, e.g., /y35 san214/: the surface word /y35 san214/ is a gap in the Mandarin lexicon, whereas the underlying word /y214 san214/ means 'umbrella.' Therefore the design of the stimuli invited lexical knowledge to play a substantial role in the task, and did not convince me that the observed mapping between surface words and underlying words was due to abstract phonological knowledge.

A. Chen and Kager (2011) and A. Chen et al. (2015) designed their studies in a way to allow the use of lexical knowledge to play a smaller role. A. Chen and Kager (2011) found that Mandarin Tone 3 and Tone 2 were discriminated by native listeners more accurately when Tone 3 preceded Tone 2, than when Tone 2 preceded Tone 3, indicating that the Tone 2 in the Tone 2 Tone 3 sequence might be perceived by the native listeners as a sandhied tone with an underlying identity of Tone 3. In this study all tones were realized on the syllable /ma/. Though /maT3 马/ 'horse' is an existing word in Mandarin, the underlying disyllabic /maT3 maT3/ does not exist in the Mandarin lexicon. In a following disyllabic tone discrimination study, A. Chen et al. (2015) found that native Mandarin listeners confused the disyllabic sequences /Tone 2 Tone 3/ and /Tone 3 Tone 3/ more than the sequences /Tone 3 Tone 2/ and /Tone 3 Tone 3/, although the words they used bearing /Tone 3 Tone 3/ are gaps in the Mandarin lexicon. Next to A. Chen and Kager (2011), the results also suggest that the native listeners may have mapped the Tone 2 in the /Tone 2 Tone 3/ string to its underlying identity Tone 3. Though these two studies adopted Mandarin characters, they avoided Peng (2000)'s use of existing disyllabic words as the underlying words and non-existing disyllabic words as the surface words. Hence lexical knowledge could not have been the reason that directly lead the listeners to map the surface words onto the underlying words in this study; abstract phonological knowledge must have played a larger role in it.

1.4.3 Summary

Based on the studies reviewed above, it may be concluded that native listeners, when hearing the sounds that have undergone contextual changes, are able to perceptually de-neutralize the alterations caused by the conditioning context, and recover the underlying sound or the underling lexical items. For assimilatory processes, a wide range of segmental and tonal studies support native listeners' ability to perform the surface-to-underlying mapping. This mapping does not necessarily rely on acoustic traces of the underlying elements, and is always context-sensitive and lexicon-independent, indicating that it may exclusively depend on the phonological knowledge about the native language. Studies on native surface-to-underlying mapping in dissimilatory processes are much fewer. Still, a few tonal studies suggest that native listeners are able to recover the underlying representations using phonological or lexical knowledge.

1.5 Surface-to-underlying mapping by naïve non-native listeners

By definition, naïve non-native listeners lack knowledge of the phonological process(es) and the lexical underlying representations of a target foreign language. As discussed in Section 1.2, assimilatory processes are motivated by ease of articulation, which is assumed to be to a large extent language-universal, whereas dissimilatory processes originate from listeners' "hyper-correction," which relies on language-specific underlying representations, or alternatively, from "motor planning" difficulties, which yields less clear predictions on the language specifity/generality of listeners' knowledge it implies. Hence, naïve non-native listeners may be expected to only access the surface-to-underlying sound mapping for assimilatory processes. What they might be doing in this mapping process is to relate a surface sound to a hypothetical underlying sound, the change being licensed via the context sound that can be assumed to have triggered the underlying-to-surface change, presumably with reference to implicit knowledge of articulatory gestures, which may be generalized from the native language, or by more language-general knowledge relating non-native sounds to general articulatory settings.

1.5.1 Non-native mapping in assimilatory processes

Some empirical evidence was found supporting the hypothesis that naïve non-native listeners can perform surface-to-underlying mapping in assimilatory processes, with diverging conclusions on whether this mapping crucially relies on gradience, i.e., the recoverability of acoustic traces of the underlying sound remaining at the surface.

Most studies found that non-native listeners were able to perform the mapping between surface and underlying forms successfully only for gradient assimilatory processes, in a context-sensitive way. For instance, Gow and Im (2004) observed that naïve English listeners performing a mapping task for a gradient Hungarian voicing assimilation process (e.g., *oros dInAstiA* ‘Russian dynasty’ becomes *oro*[z] *dInAstiA*) were facilitated in monitoring a target (underlying) sound in its surface form in a viable context, which licensed the assimilatory change, as compared to an unviable context, whereas naïve English listeners did not show such facilitating effects of viable context in a Korean categorical labial-to-velar assimilation process (e.g., /gom+gæʃi/ ‘bear-like’ becomes [goŋgæʃi]). Mitterer et al. (2013) similarly observed for the Korean categorical labial-to-velar assimilation process that naïve Dutch and English groups of listeners both failed to categorize the derived velar as a target (underlying) labial more in the viable context. Instead, they showed a distinctively different pattern from the native Korean listeners, that is, they identified sounds more frequently as velar sounds before a velar context than before a control context. An ERP study (Mitterer, Csépe, Honbolygo, & Blomert, 2006) tested naïve Dutch listeners on a Hungarian liquid assimilation process (/l/ to /r/), and found that the listeners accepted the assimilated segment as its underlying counterpart more often in a viable context than in an unviable context when they heard a partially assimilated /r/, whereas they did not show the same contextual effect when they heard a categorically assimilated /r/.

More direct evidence that mappings by naïve non-native listeners rely on phonetic detail was found in a continuum study by Mitterer, Csépe, and Blomert (2006). Recall that in Mitterer et al. (2013)’s study on Korean listeners’ native mapping, the listeners interpreted only the most assimilated velar-velar sequences at the velar endpoint of the continuum as possible underlying labial-velar sequences (see Section 1.5.1). Using the same methodology, Mitterer, Csépe, and Blomert (2006) tested naïve Dutch listeners on a Hungarian liquid assimilation process (e.g., /knɑl+ro:t/ ‘vivid red’ becomes [knɑrro:t], an application of Hungarian liquid assimilation to Dutch words). They created a /bɔl/ – /bɔr/ continuum and presented it in viable following contexts such as /ro:l/ and in unviable contexts such as /na:l/, and asked listeners to identify each step of the continuum as /l/ or /r/. The listeners demonstrated a blurred distinction between the underlying /lr/ and the surface /rr/ throughout the continuum, with a slightly lower /l/ response rate towards the /r/ endpoint of the continuum. These results suggested that the naïve Dutch listeners probably mainly exploited the residual phonetic detail for the underlying /l/ which is left in the assimilated form to perform the mapping between the surface and the underlying sounds in this assimilatory process, based on general knowledge of articulatory ease.

A single study suggests that naïve non-native listeners are able to relate the surface and the target (underlying) sounds in categorical assimilation processes. Darcy et al. (2009) tested naïve French listeners on a categorical English place assimilation process as well as naïve English listeners on a categorical French voicing assimilation process, using a word detection experiment (details of this paradigm will be presented in Chapter 4), and observed that both groups of listeners detected the target underlying word more successfully in viable than in unviable contexts, though neutralized phonemic contrasts were used in both the processes. This effect was small but statistically significant. Taken together the different results in the above studies, it still needs to be confirmed whether or not the mapping between surface and underlying sounds in assimilatory processes for non-native listeners is restricted to gradient changes.

1.5.2 Dissimilatory processes

Regarding dissimilatory processes, no previous study has investigated whether naïve non-native listeners can perform a mapping between surface and underlying sounds, as far as I am aware. Hence, it remains unknown whether the surface-to-underlying mapping in dissimilatory processes is inaccessible for naïve non-native listeners, as the “hyper-correction” theory (Ohala, 1993) implied. One of the goals of this dissertation is to fill this void, in order to understand how the language-specific and language-general knowledge of native and naïve non-native listeners operate in the mapping problem.

1.5.3 Summary

This section discussed naïve non-native listeners' capacity for surface-to-underlying sound mapping in assimilatory and dissimilatory processes. Some studies observed this mapping in assimilatory processes, most of which involved gradient changes that left fine-grained phonetic traces for underlying segments in assimilated forms, suggesting that the mapping seems to depend largely on these remaining articulatory gestures for the underlying segment. This finding is compatible with the language-general mapping hypothesis for assimilatory processes. When it comes to dissimilatory processes, a gap remains in the experimental literature regarding whether the surface-to-underlying mapping in these processes is inaccessible for naïve non-native listeners. In order to test the language-specific mapping hypothesis for dissimilatory processes, it is important to fill this gap.

1.6 Research questions and dissertation outline

To recapitulate, the general research questions I pursue in this dissertation are:

> **Research question 1**: Is listeners' surface-to-underlying mapping dependent exclusively on language-specific knowledge, or is it also facilitated by language-general articulatory knowledge (comparing native and non-native listeners)?
>
> **Research question 2**: Will the mapping difficulty be different between assimilatory and dissimilatory processes, especially for naïve non-native listeners (assimilatory vs. dissimilatory)?
>
> **Research question 3**: Is the mapping further influenced by the categoricalness/gradience of assimilatory and dissimilatory processes, for naïve non-native listeners (categorical vs. gradient)?

I decide to examine these three aspects of the mapping issue in tone and tone sandhi phenomena, because non-tonal language listeners are maximally distant from native tonal language listeners, because they do not have previous exposure to tones, hence guaranteed to have no experience with tone or tone sandhi (as argued above in Section 1.1), and are thus suitable for testing the effect of language-general articulatory knowledge.

Chapter 2 will be a production study on two pairs of comparable assimilatory and dissimilatory tone sandhi processes in Nanjing Mandarin that are documented in the literature. This chapter will be outlined after Chapter 3 and 4, because Chapter 3 and 4 are the mapping studies that address the general research questions 1-3. Note that the outline of Chapter 3 will provide the motivation for conducting the production study in Chapter 2 in the first place.

The first mapping study, reported in Chapter 3, will deal with the general research questions 1 and 2. In this study, Nanjing Mandarin, a branch of Jianghuai Mandarin spoken in the Yangtze area of China, is selected as the target tonal language. As recorded in previous studies (Jiangsusheng Difangzhi Bianzuan Weiyuanhui (Editorial Committee of Jiangsu Province Chorography), 1998; Liu, 1995; Song, 2006), Nanjing Mandarin has two pairs of comparable assimilatory and dissimilatory tone sandhi processes, each pair involving a fixed tonal alternation between an underlying tone and a surface tone. Nanjing Mandarin listeners will be the native tonal language group in this mapping study, and Dutch listeners will be the completely naïve non-tonal language group. They will participate in a Concept Formation experiment (Jaeger, 1986). In this experiment, participants in each language group are randomly assigned to the assimilatory or dissimilatory Nanjing sandhi process condition, and then further randomly divided to be trained on the underlying or surface tone of the process as the target category. After the training, they are tested on the sandhied tone in the tonal context that licences the sandhi process, to see if they could perform the surface-to-underlying tone mapping in their own way. This study will address the following specific research questions:

1) Will native Nanjing Mandarin listeners be the only group to successfully perform surface-to-underlying tone mapping in Nanjing Mandarin sandhi processes, or will Dutch listeners also be able to perform the mapping, based on language-general articulatory knowledge? (This research question is an instantiation of the general research question 1.)

2) Will assimilatory tone sandhi processes (compared to dissimilatory processes) lead to facilitated surface-to-underlying mapping for naïve Dutch listeners? (This research question is an instantiation of the general research question 2.)

A second mapping study, presented in Chapter 4, will focus on surface-to-underlying mapping by naïve non-native listeners. It will continue to investigate the effect of the assimilatory vs. dissimilatory processes in non-native mapping, using a cognitively less challenging task: a Word Detection experiment (based on the observation that the task used in Chapter 3 proved to be challenging for the Dutch listeners). In addition, this study will address the dissertation's research question 3 concerning categorical vs. gradient processes, to see if the hypothesized facilitated mapping in assimilatory processes is only observed when the process is gradient. This study will adopt a set of artificial tone sandhi rules, to allow a more controlled tone alteration (as compared to that in Chapter 3) from the underlying tone to the surface tone. Once again, Dutch listeners will serve as the naïve non-tonal language group. In this experiment, naïve Dutch listeners will be asked to detect the target words in the underlying tonal shape of the assimilatory/dissimilatory sandhi processes in surface items containing the target words in the sandhied tonal shape of the processes. One specific research question that this study aims to answer is identical to the research question 2 in Chapter 3 (by using a different methodology). A second research questions that it addresses specifically is:

3) Will gradient and categorical processes further lead to any difference in surface-to-underlying mapping for Dutch listeners? In specific, will Dutch listeners rely exclusively on the residual phonetic details for the underlying tone which are left in the surface tone, in order to perform the surface-to-underlying mapping in Nanjing assimilatory sandhi rules? (This research question is an instantiation of the general research question 3.)

Preceding the above two mapping studies, this dissertation will include a production study in Chapter 2 on the aforementioned two pairs of comparable assimilatory and dissimilatory tone sandhi processes in Nanjing Mandarin, in order to verify earlier studies and to decide if any of these processes can be used as the basis for creating stimuli in the Nanjing Mandarin mapping study of Chapter 3. In this study, native speakers of Nanjing Mandarin will be recorded in a disyllabic tone sandhi elicitation task to produce natural disyllabic tone sandhi patterns; the productions will be acoustically analysed, and then a pair of assimilatory and dissimilatory tone sandhi processes will be selected to be used in the mapping study of Chapter 3. This study will, as a stand-alone production study, investigate if assimilatory and dissimilatory tone sandhi rules will result in different tonal acoustic realizations in native speakers, in terms of *categorical versus gradient* application of tone sandhi, i.e., categorical when an underlying tone becomes neutralized with another tone in the surface form, or gradient, when pitch features of an underlying tone are only partially lost from the surface form.

As explained, this dissertation has three overarching general research questions, each related to specific research questions in Chapter 2-4, which will be written in the style of individual journal articles. The research questions of Chapter 3 and 4 are developed from the dissertation's general research questions concerning surface-to-underlying mapping; the research questions of Chapter 2 are stated to examine one key effect in the mapping issue - assimilatory and dissimilatory processes, from a production perspective. There may be some repetitive background literature in these chapters and also in Chapter 1.

Finally, Chapter 5 will conclude all the major findings in the three experimental studies, and present general discussions.

Chapter 2

An acoustic comparison between assimilatory and dissimilatory tone sandhi processes in Nanjing Mandarin：The role of categoricalness versus gradience[1]

2.1 Introduction

In tonal languages, lexical tones may undergo alterations under the influence of their contextual tones. This phenomenon is generally known as *tone sandhi* (M. Y. Chen, 2000). Tone sandhi processes can be classified as “assimilatory” and “dissimilatory” based on whether the altered tone becomes more or less similar to its context tone. When the sandhied tone approaches the context tone in pitch (f0), the tone sandhi process is defined as *assimilatory*. For example, Zhenjiang Mandarin (Qiu, 2012) has an assimilatory tone sandhi rule, which turns a high falling tone immediately before another high falling tone into a high level tone, with tone offset raised to the same pitch height as the following tone’s onset,

[1] Preliminary results of this chapter were presented at *the 11th International Symposium on Chinese Spoken Language Processing,* and published in the proceedings as X. Li and Kager (2018a).

written as HL.HL → HH.HL. (Each tone is represented by the pitch features (H/L) of its onset and offset; the dot is used to separate pitch features of the tones.) When the sandhied tone deviates from the context tone in pitch (f0), while they underlyingly share the same pitch features, the tone sandhi process is referred to as *dissimilatory*. For example, The T3 sandhi rule in Beijing Mandarin is an example of the dissimilatory tone sandhi process (Cheng, 1973; Yip, 1980, 2002): a low T3[2] immediately preceding another low T3 changes to a low-rising T2, transcribed as LL.LL → LH.LL (e.g., Speer et al., 1989; Yip, 1980), with the altered tone offset deviating from the following tone's onset.

Assimilatory and dissimilatory processes are widely believed to have distinct origins, the former grounded in articulatory ease (e.g., Grammont, 1933; Johnson, 1973) and the latter not, as will be explained later (e.g., Ohala, 1993). The general question we pursue in this study is whether the different motivations for assimilatory and dissimilatory tone sandhi processes result in different acoustic realizations in native speakers, in terms of *categoricalness* versus *gradience.* A tone sandhi rule is *categorical* when the derived tone becomes neutralized with another tone, with pitch features of the underlying tone lost from the surface form; whereas a sandhi rule is *gradient* when pitch features of the underlying tone are only partially lost from the surface form and hence no neutralization takes place.

Previous experimental studies have looked into the categoricalness/gradience of assimilatory and dissimilatory tone sandhi processes in native speakers in many tonal languages/dialects. For example, Myers and Tsay (2008) found that in Taiwan Southern Min a high level *yinping* (in traditional tone classification of this dialect) is categorically assimilated and neutralized to a mid level *yangqu* before a low rising *yangping* (55.24 → 33.24[3]); the above-mentioned dissimilatory Beijing Mandarin T3 sandhi rule (LL.LL → LH.LL) was observed to involve a non-neutralizing change from underlying low T3 to rising surface T2 before

[2] The distinctive feature of lexical Mandarin T3 is "low" (e.g., Speer, Shih, & Slowiaczek, 1989; Yip, 1980).

[3] The numbers indicate tone values on Chao (1930)'s five-level scale, with 1 representing the lowest pitch level in the speaker's pitch range, and 5 representing the highest; the dot is used to separate tones.

another low T3[4] (Myers & Tsay, 2003; Peng, 2000; Yuan & Chen, 2014). However, no general conclusions can be drawn from comparing these studies about the relationship between assimilatory/dissimilatory tone sandhi on the one hand and categoricalness/gradience of sandhi application on the other, as cross-linguistic comparisons may be hampered by uncontrolled language-specific factors such as phonological processes or sound inventories. Also, the experimental and analytical methodologies adopted in these studies vary from each other. More specifically, the studies differ in their criteria regarding what constitutes categoricalness of tone sandhi application.

A single-language setting is crucial to making a fair comparison between assimilatory and dissimilatory processes. A few studies targeting neutralization in specific tonal languages/dialects involved a number of tone sandhi processes. However, the acoustic data and the phonological notations of the tone sandhi processes in these studies often do not fully agree with each other, and hence whether these tone sandhi processes are assimilatory or dissimilatory in nature is not always clear. For example, Jie Zhang and Liu (2011) reported a categorical tone sandhi rule which changes a mid falling T1 before another T1 to a neutralized high rising T2, with the onset of the second T1 raised as well due to assimilatory carryover effect. This process was transcribed as L.L → LH.L (hence, dissimilatory) in their study. Yet it may also be interpreted as an assimilatory process, because the offset of the first T1 is raised to the same pitch (f0) height as the onset of the following T1. In the same study, they also observed a gradient tone sandhi process which changes a low dipping-rising T3 preceding another T3 to a tone resembling a lexical mid-rising T2 but lower in overall mean pitch (f0). This process seems to be dissimilatory, similar to the aforementioned Tone 3 sandhi rule in Beijing Mandarin, because the offset of the first T3 deviates far from the onset of the following T3. Yet it was transcribed in their study as LH.LH→ H.LH, showing dissimilation with respect to the abstract phonological rule that describes a change of the first tone LH in a sequence of two identical tones, but fails to show dissimilation in terms of acoustic values of tone onsets and offsets.

---

[4] The sandhied T3 was found to be perceptually indistinguishable from an underlying T2 for native listeners (Peng, 1996; Speer et al., 1989; Wang & Li, 1967).

No earlier study has compared the categoricalness/gradience of undisputed assimilatory and dissimilatory processes within a single tonal language, as far as we are aware. Hence, a study of this type is needed to establish whether there are differences in tone sandhi categoricalness/gradience between carefully selected pair(s) of tone sandhi processes that differ in their assimilatory/dissimilatory nature, but are otherwise maximally comparable in terms of f0 trajectories.

Predictions can be made regarding the categoricalness/gradience of assimilatory and dissimilatory processes from the different origins of these processes. Assimilatory processes are mostly believed to be motivated by increasing ease of articulation (e.g., Grammont, 1933; Johnson, 1973). Kiparsky (1995) made the prediction that assimilatory processes tend to start out as gradient, because sound changes due to articulatory processes are inherently gradient, and the changes become reinterpreted as phonological rules by listeners in a slow diachronic progress.

With regard to dissimilatory processes, the most widely accepted view concerning their motivation is Ohala (1993)'s proposal that a dissimilatory sound change occurs because listeners erroneously recover a speaker's intended pronunciation upon hearing a surface signal, misattributing it to an underlying sound which (by definition) instantiates an existing phonological category in the language (termed "hyper-correction"). Therefore it predicts dissimilatory processes to always be categorical, as claimed explicitly by Ohala (Ohala, 1987, p. 218; 1993, p. 255-256) and by later researchers in their various discussions of the "hyper-correction" theory (Garrett, 2015, p. 236; Jatteau & Hejná, 2016, p. 361; Kiparsky, 1995, p. 658).

An alternative classic view on the motivation of dissimilatory processes is the "motor planning" account (Frisch, 2004; Frisch et al., 2004; Garrett, 2015; Garrett & Johnson, 2012; Grammont, 1895, 1933; Tilsen, 2007). "Motor planning" is generally viewed as the process of "constructing or retrieving motor plans that will later be executed by speaking" (Garrett & Johnson, 2012, p. 59). According to Garrett and Johnson (2012), this view was first proposed by Grammont (Grammont, 1895, 1933), who argued (as cited in Garrett and Johnson (2012, p. 57)) that non-local dissimilation occurs "when planning for a segment in a more prominent position distracts a talker who is producing a similar segment in a weaker position.” For example, Frisch (2004, p. 346) and Frisch et al. (2004, p. 180) proposed that dissimilation may occur because of the difficulty of processing identical or similar sounds, a scenario interpreted by Garrett and Johnson (2012) as belonging to “motor planning.” Perhaps the most insightful discussion of “motor planning” as the motivation for dissimilation was offered by Garrett (Garrett, 2015; Garrett & Johnson, 2012). Garrett and Johnson (2012) proposed that (non-local) dissimilation is caused by the speakers “inhibition of one segment by the activation of another” due to the preference of an alternating sound pattern over a repetitive pattern (p. 60). Based on this account, Garrett made predictions in subsequent papers regarding the categoricalness/gradience of dissimilation changes, which upon closer inspection turn out to be somewhat inconsistent. Garrett and Johnson (2012) stated explicitly that “sound changes rooted in motor planning and perceptual parsing are often categorical” (p. 78). This expectation is relevant for (non-local) dissimilation. However, in a later study (Garrett, 2015) it was observed that “motor planning” only accounted for an early stage of sound change in which one of two identical gestures is only reduced instead of lost (p. 236), implying that dissimilatory processes can be gradient. This observation was taken by Jatteau and Hejná (2016) as Garrett’s prediction for the categoricalness/gradience of dissimilatory processes. However, we believe that this interpretation may involve an over-extension of Garrett’s (2015) observation because of the different prediction made in Garrett and Johnson (2012), as argued above.

Tilsen (2007, 2013) showed that “motor planning” inhibition as the motivation for dissimilation is supported by empirical evidence. He observed in a primed vowel-shadowing task that when participants were asked to shadow an /i/ after a prime /a/ and a target /i/, a majority of participants tended to produce an /a/ instead (Tilsen, 2007). Similar results were found in a primed tone-shadowing task (Tilsen, 2013) in which participants tended to produce the prime tone (which occurred before the target tone) when being asked to shadow the target tone. He interpreted these results as evidence suggesting that dissimilation arises from motor planning inhibition. Tilsen (2013) is the only study we find in the "hyper-correction" and "motor planning" literature that specifically discusses tone dissimilation.

Taking together the studies on "motor planning" as motivation for dissimilation, it is not totally clear whether this account predicts exclusively categorical dissimilatory processes or allows both categorical and gradient processes.

The specific goal of this study is to select pairs of comparable assimilatory and dissimilatory tone sandhi processes from a single tonal language, and look into the categoricalness/gradience of the tone sandhi applications by native speakers in order to test a) the prediction that assimilatory processes can be gradient and b) the different predictions made by the "hyper-correction" and the "motor planning" views regarding dissimilatory processes. The "hyper-correction" view would predict that dissimilatory processes should always be categorical, hence never show any gradient application. As stated above, predictions from the "motor planning" account are less clear. Nevertheless, if we observe gradient dissimilatory processes in this study, it will run against the "hyper-correction" view.

A tone sandhi process will be judged in this study as categorical or gradient based on an acoustic comparison between the sandhied tone and its claimed target surface tone in the literature. For example, if a tone sandhi process is documented in the literature as T3.T1 → T2.T1, in the current study the sandhied form of T3 in T3.T1 will be acoustically compared against the target surface T2 in T2.T1. To be more specific, the sandhi process will be judged as gradient in this study if an acoustic difference is observed between the sandhied tone and the target surface tone and the sandhied tone forms an incomplete change towards the claimed target surface tone. The process will be judged as categorical if a complete change of the sandhied tone occurs towards the claimed target surface and no acoustic difference is observed between the sandhied tone and the claimed target surface. In addition, speaker variation will be considered in order to assess whether the observed acoustic difference is a true manifestation of the gradient nature of the sandhi process or, alternatively, is due to speaker variation. For example, if some speakers apply tone sandhi in a categorical fashion whereas other speakers do not apply it at all, summation over speakers would result in an incorrect judgment of gradience.

The current study will investigate the categoricalness/gradience of assimilatory and dissimilatory tone sandhi processes in Nanjing Mandarin, a tonal language of China known to have both assimilatory and dissimilatory tone sandhi processes (Jiangsusheng Difangzhi Bianzuan Weiyuanhui (Editorial Committee of Jiangsu Province Chorography) [JDBW], 1998; Liu, 1995; Song, 2006). Nanjing city is located in the southwest of Jiangsu Province and is its provincial capital. Nanjing Mandarin, a representative dialect in the Hongchao dialectal area of Jianghuai Mandarin, is a branch of Northern Mandarin in China.

2.2 Nanjing Mandarin lexical tones and disyllabic tone sandhi patterns documented in literature

There have been a number of experimental studies on lexical tones and disyllabic tone sandhi patterns of Nanjing Mandarin, of which the most cited are Liu (1995), JDBW (1998), and Song (2006). Liu (1995) and JDBW (1998) documented trained dialectologists' impressionistic transcriptions on speakers aged between 60 and 90 years old. Song (2006) used experimental means to document the tone patterns from 2 native speakers aged between 50 and 60. These studies all used a small group of older speakers, and no recent work has been done on disyllabic tone patterns of Nanjing Mandarin. The lexical tones and disyllabic tone sandhi rules they derived are described as below.

2.2.1 Lexical tones

Nanjing Mandarin has five distinctive lexical tones (neutral tone excluded). Table 1 summarizes the transcriptions of the tones recorded in Liu (1995), JDBW (1998), and Song (2006). T1 in Nanjing Mandarin is recorded in these studies as a high/mid-falling tone; T2 is a low-rising tone; T3 is recorded as a low level/dipping tone; T4 is a high/mid level tone; T5 is a high and abrupt tone.

Table 1. *Transcriptions of the five lexical tones of Nanjing Mandarin in previous studies. (A single number in the tone value indicates a short and abrupt tone.)*

Reference	T1	T2	T3	T4	T5
Liu (1995)	41	24	11	44	5
JDBW (1998)	31	24	11	44	5
Song (2006)	41	24	22/212	44/33	5

2.2.2 Disyllabic tone sandhi patterns

Liu (1995), JDBW (1998), and Song (2006) all agreed on six disyllabic tone sequences that undergo tone sandhi, though their transcriptions for the lexical tones did not always agree with each other, as shown in Table 2. From their transcriptions of the sandhi rules, we observe that T1.T1 and T3.T1 undergo assimilatory tonal changes, and T2.T5, T3.T3, T4.T5, and T5.T5 undergo dissimilatory tonal changes. We can assume that four of the tone sandhi processes are agreed in these studies to involve a categorical change to another existing tone: 1) T1.T1 → T4.T1, 2) T2.T5 → T3.T5, 3) T3.T1 → T2.T1, and 4) T3.T3 → T2.T3, though their categoricalness needs to be examined with new data. It was controversial in the studies whether T4.T5 changes to T1.T5 in a categorical way: Liu (1995) documented so, but JDBW (1998) and Song (2006) thought this tonal change was gradient.

Table 2. *Transcriptions of the disyllabic tone sandhi rules in Nanjing Mandarin in previous studies.*

	Liu (1995)	JDBW (1998)	Song (2006)
T1.T1	41.41 → 44.41	31.31 → 44.31	41.41 → 44.41
T2.T5	24.5 → 11.5	24.5 → 22.5	24.5 → 22.45
T3.T1	11.41 → 24.41	22.31 → 24.31	212.41 → 24.41
T3.T3	11.11 → 24.11	22.22 → 24.22	212.212 → 24.212
T4.T5	44.5 → 41.5	44.5 → 43.5	44.5 → 43.45
T5.T5	5.5 → 3.5	5.5 → 4.5	5.5 → 53.5

2.2.3 Two pairs of comparable assimilatory and dissimilatory tone sandhi processes

Consulting the above literature on Nanjing Mandarin tone sandhi patterns, we found two pairs of comparable assimilatory and dissimilatory tone sandhi processes. The first sandhi pair is composed of assimilatory Sandhi 1[5] and dissimilatory Sandhi 2, both of which involve a tonal change between T1 and T4. In assimilatory Sandhi 1 (T1.T1 → T4.T1), the offset of a high-falling T1 in the first syllable is raised to the same pitch as the onset of the following T1. This sandhi process was transcribed as HL.HL → HH.HL in Ma (2009), which gave symbolic transcriptions for the sandhi process in a way consistent with Liu (1995), JDBW (1998), and Song (2006)'s phonetic transcriptions in Table 2. This sandhi process changes a high falling T1 preceding another T1 to a high level tone resembling a lexical T4. In dissimilatory Sandhi 2 (T4.T5 → T1.T5), the offset of a high-level T4 in the first syllable deviates from the high onset of the following T5 instead of approaching it, transcribed as HH.H → HL.H (Ma, 2009). This sandhi process changes a high level T4 preceding a high T5 to a high falling tone resembling a lexical T1. Though the underlying-to-surface tonal changes are opposite in direction in the two sandhi processes, both sandhi processes involve a T1 vs. T4 contrast, with equal distance (in the sense of tone categories) between the underlying tone and the surface tone. Thus, these constitute a pair of assimilatory and dissimilatory tone sandhi processes that are comparable with each other.

The second sandhi pair is composed of assimilatory Sandhi 3 and dissimilatory Sandhi 4, both of which involve a tonal alternation between an underlying T3 and a surface T2, and hence constitute another comparable pair of assimilatory and dissimilatory tone sandhi processes. Assimilatory Sandhi 3 (T3.T1 →T2.T1) changes a low level/dipping T3 preceding a high falling T1 to a low rising T2, transcribed as LL.HL → LH.HL (Ma, 2009). Dissimilatory Sandhi 4 (T3.T3 → T2.T3) changes the low level/dipping T3 preceding another T3 also to a low rising T2, transcribed as LL.LL → LH.LL (Ma, 2009).

---

[5] Sandhi 1, Sandhi 2, Sandhi 3, Sandhi 4, and Sandhi Pair 1&2 are labeled by the authors of this study.

In sum, this study investigates the categoricalness/gradience of Sandhi Pair 1: assimilatory Sandhi 1 versus dissimilatory Sandhi 2 and Sandhi Pair 2: assimilatory Sandhi 3 versus dissimilatory Sandhi 4 (as summarized in Table 3), collecting new data from native Nanjing Mandarin speakers.

Table 3. *Two pairs of comparable assimilatory and dissimilatory tone sandhi processes in Nanjing Mandarin.*

	Sandhi Pair 1		Sandhi Pair 2	
	Sandhi 1	Sandhi 2	Sandhi 3	Sandhi 4
Underlying form	T1.T1	T4.T5	T3.T1	T3.T3
Surface form based on the literature	T4.T1	T1.T5	T2.T1	T2.T3
	Assimilatory	Dissimilatory	Assimilatory	Dissimilatory

2.3 Methodology

In this study, we investigated the lexical tones and disyllabic tone sandhi processes of Nanjing Mandarin produced by young people aged 18 to 30 year old, in contrast with the older age groups adopted in previous studies. This study also invited a larger number of speakers as compared to previous studies to produce more reliable results.

2.3.1 Participants

Eighteen native speakers of Nanjing Mandarin (9 females and 9 males) were recruited for the experiment. All speakers were aged between 18 and 30 years old. They were born and raised in Nanjing and had never left Nanjing for more than half a year. None of them had any speaking or hearing problems. They were financially compensated for the experiment.

2.3.2 Stimuli

For each of the five lexical tones of Nanjing Mandarin, four monosyllabic words were selected. The monosyllabic words were all frequently used and easy-to-combine Chinese characters used in Nanjing Mandarin. The distribution of front-back and high-low vowels was balanced across the words. Since these monosyllabic words were to be used to combine disyllabic words with other monosyllabic words, zero consonants were excluded as much as possible for the convenience of splitting syllables. For each monosyllabic word Ta (T1/T2/T3/T4/T5), there were 10 disyllabic combinations, namely, 5 disyllabic combinations of Ta.Tx ($_X$ =T1, T2, T3, T4, T5) when Ta was in the first syllable, and 5 disyllabic combinations of Tx.Ta when Ta was in the second syllable. Therefore, the word list contained 220 words (11 tone types (1 monosyllabic word + 10 disyllabic sequences) * 4 monosyllabic words * 5 lexical tones) (for the full word list, see Appendix A).

Most of the disyllabic words in the word list were selected from Liu (1995), which documented the typical Nanjing Mandarin vocabulary in the 1990s. A few other words were common Mandarin items that were also used in the Nanjing Mandarin area. All the test words were examined by a native female Nanjing Mandarin speaker living in Utrecht, The Netherlands, and all of the words were reported to be common words. All of the words in the word list were then randomized to form the stimulus list.

2.3.3 Recording

Recordings of all the 18 Nanjing Mandarin speakers took place in a quiet room in Nanjing University in Nanjing, China. Recordings were conducted with a headset Philips microphone using the software Cool Edit Pro on a MacBook laptop with 44.1 kHz sampling rate and 16 bit rate in mono.

The stimuli were presented one by one in simplified Chinese characters on the laptop screen. Speakers were seated in front of the screen and read each stimuli two times naturally in Nanjing Mandarin. They were able to control the speed of stimulus presentation, but generally paused 1-2 seconds between trials. They could take a rest halfway when they felt needed, and repeating was required when they misread or read too quickly. Altogether 40 monosyllabic tokens and 400 disyllabic tokens were elicited from each of the 18 participants.

2.3.4 F0 analysis

The segmentation was manually conducted using Praat 4.3.09 (Boersma & Weenink, 2009). The f0 values of the tone contour of each word were measured at 10 equidistant points (P1…P10) in the tone-bearing rhyme part of the syllable, resulting in a set of time-normalized f0 values, which eliminated the difference among the durations of every rhyme part to get f0 curves that could be compared with each other. These values were automatically extracted by a self-written script in Praat and saved together with their original time scale. Obvious errors of the f0 algorithm (e.g., octave jumps) were corrected manually.

In order to eliminate the anatomical variation between speakers caused by their different sizes of vocal folds and to allow between-speaker comparison, raw Hertz values were transformed to semitones relative to each speaker's average pitch (f0) (Andruski & Costello, 2004).

2.4 Results

2.4.1 Lexical tones produced in isolation

Figure 1 shows the 5 lexical tones of Nanjing Mandarin averaged across the 18 speakers (9 males and 9 females). As can be seen from Figure 1, T1 is a high-falling tone with a pitch contour that falls from the higher to the lower end of the speaker's pitch range; T2 is a low-rising tone, whose pitch contour rises from the lower to the upper end of the pitch range; T3 is a low falling tone, which falls slightly from the lower pitch range to even lower; T4 is a mid-high level tone which stays at the mid-high level of the pitch range; T5 is a salient high tone with a concave contour shape, and it is short and abrupt with respect to its duration. The tonal shapes of the lexical T1, T2, T4, and T5 observed in our study mostly replicate those in previous studies, yet we find the overall falling tendency in lexical T3, which may be a new feature of the productions by the younger age group.

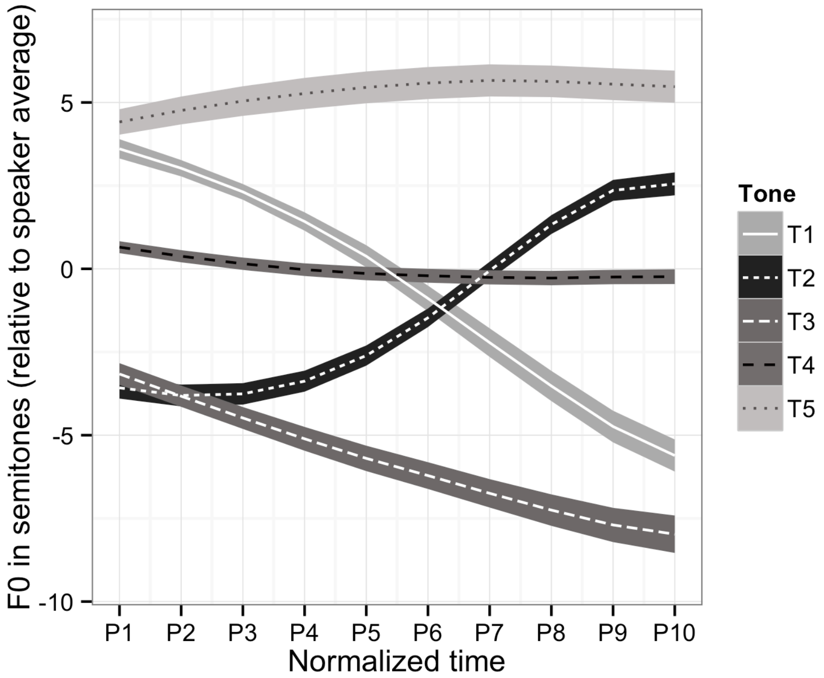

Figure 1. *Five lexical tones in Nanjing Mandarin produced in isolation (with normalized time) by 18 speakers. Central lines represent the means. Shaded areas (ribbons) stand for ±1 standard error of the mean.*

2.4.2 Categoricalness/gradience of the two pairs of assimilatory and dissimilatory sandhi processes

To compare the sandhied tones against their claimed target surface tones, growth curve analysis (Mirman, 2014) was employed in R (R Core Team, 2014) using the *lmer* function in the *lme4* package (Bates, Maechler, Bolker, & Walker, 2014). The advantage of this analysis method is that it evaluates the overall f0 contours of the tones. It was successfully used by Q. Li and Chen (2016) and Jie Zhang and Meng (2016) to determine whether there is tone neutralization in Tianjin

Mandarin and Shanghai Wu tone sandhi processes, respectively. In growth curve analysis, a non-linear f0 curve is fitted with the formula in (1), with *x* representing time, *y* representing f0, and different coefficients indicating different features in f0 curves. The intercept *a* indicates the overall f0 mean of a curve; *b* indicates the slope of the f0 change; *c* represents the sharpness of the centered peak, etc. (Jie Zhang & Meng, 2016). We follow Q. Li and Chen (2016) and Jie Zhang and Meng (2016) to include up to second-order polynomials to model all the tones in the current study, in order to adequately model for f0 contours of the tones and avoid overfitting.

(1) $y = a + bx + cx^2 + dx^3 + ex^4 ...$

The dependent variable for the models is f0 converted to semitone values relative to each speaker's average f0. Each model was first built with only the fixed effects of linear and quadratic time terms, and the random effects of speaker and speaker-by-tone on the time terms (base model). Then the fixed effects of tone on the intercept, the linear time term, and the quadratic time term were added in the model in a stepwise way. The significance of each effect was evaluated by log-likelihood model fit comparison at each step. A significant fixed effect of tone on the intercept, on the linear time term and on the quadratic time term would suggest that the sandhied (surface) tone and the claimed surface tone are different in overall f0 mean, f0 slope, and sharpness of the centered peak, respectively.

2.4.2.1 Sandhi Pair 1: assimilatory Sandhi 1 vs. dissimilatory Sandhi 2

2.4.2.1.1 Categoricalness/gradience of Sandhi 1

Figure 2 depicts the sandhied form of T1.T1 against its claimed target surface tone pattern T4.T1, averaged across 18 speakers. The growth curve analysis revealed that the sandhied T1 on syllable 1 (T1(syl1)) in T1.T1 and the T4 in T4.T1 are marginally significantly different in overall f0 mean (Estimate = -0.31, t = -2.43, $p < .05$ *). However, they are not significantly different in f0 slope (Estimate = 0.23, t = 0.99, $p > .05$) or sharpness of the centered peak (Estimate = -0.17, t = -1.41, $p > .05$). It seems that the sandhied T1(syl1) in T1.T1 is near-neutralizing with its claimed surface tone T4 in T4.T1, which suggests that assimilatory Sandhi 1 applies in a near-categorical way.

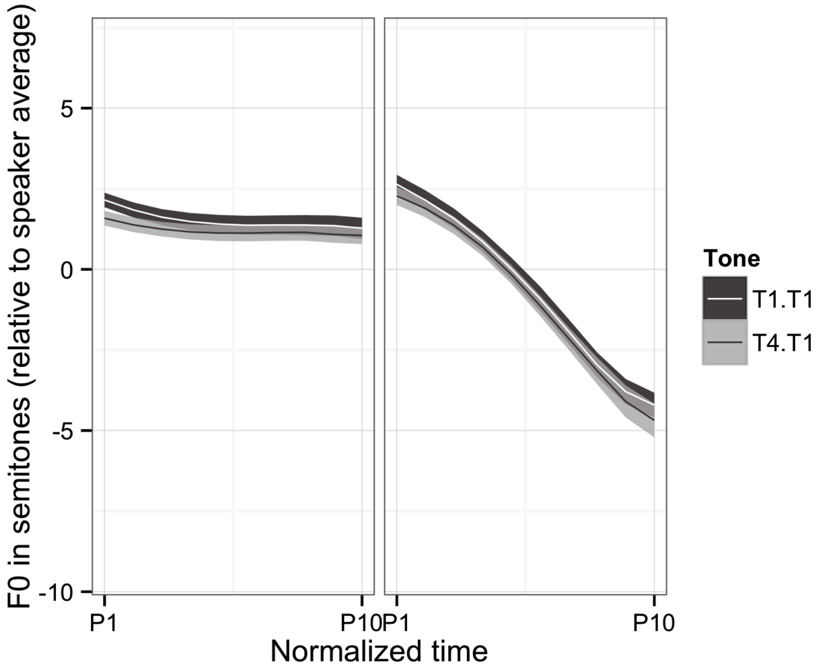

Figure 2. *Sandhied T1.T1 vs. its claimed surface T4.T1 in Nanjing Mandarin (with normalized time) by 18 speakers. The lefthand panel shows the pitch contours (in semitones relative to speaker average) of the first syllables; the righthand panel shows those of the second syllables. Central lines represent the means. Shaded areas (ribbons) stand for ±1 standard error of the mean.*

The random effect (intercept, linear, and quadratic) estimates of the T1(syl1) in T1.T1 and the T4 in T4.T1 for each speaker were extracted from the best-fitting model (for a detailed report on the estimates, see Appendix B). The differences of the estimates between the two tones were calculated for each speaker, and then standard deviations for the intercept, linear and quadratic differences across the speakers were computed. The standard deviation for the intercept, linear, and quadratic difference between the T4 in T4.T1 and the T1(syl1) in T1.T1 across speakers is 0.74, 0.84, and 0.18 (rounded off to 2 decimals), respectively, which suggests speaker variation in overall f0 mean, f0 slope, and sharpness of the centered peak differences between the two tones. Three speakers (speakers 2, 11, and 14) evidently (from Figure 3) applied the sandhi process in a gradient way, with the slope of the sandhied tone steeper than a lexical T4 (target surface); most of the speakers (speakers 4, 6, 12, 13, 15, 16, etc.) applied the tone sandhi process in a categorical way; none of the speakers did not apply sandhi at all.

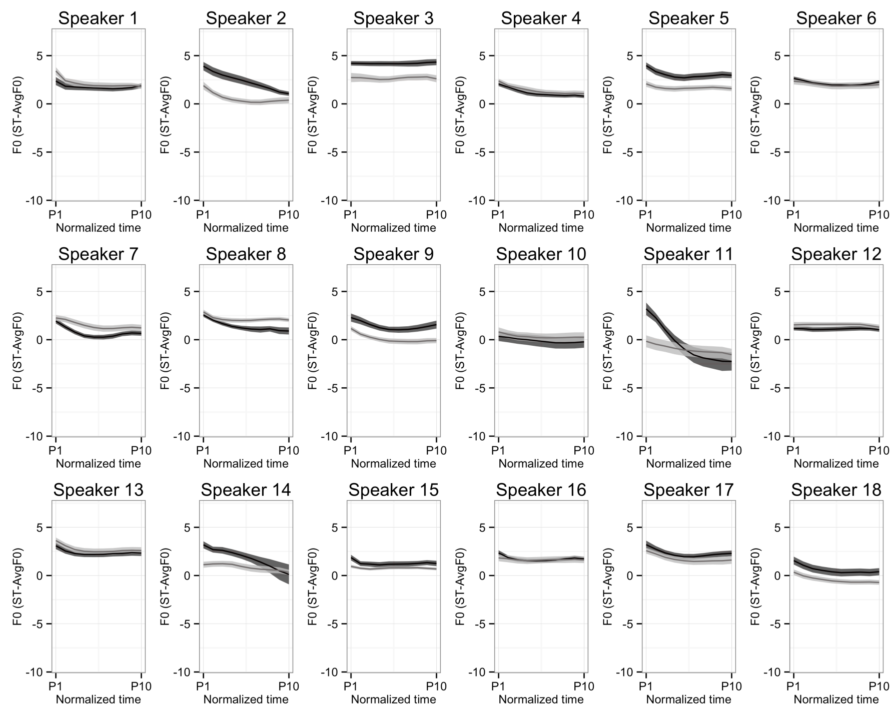

Figure 3. *T1(syl 1) in T1.T1 and T4 in T4.T1 produced by different speakers (the darker shades stand for the T1(syl 1) in T1.T1, the lighter shades stand for the T4 in T4.T1). ST-AvgF0: semitone relative to speaker average.*

2.4.2.1.2 Categoricalness/gradience of Sandhi 2

Figure 4 depicts the sandhied form of T4.T5 against its claimed target surface tone pattern T1.T5, averaged across 18 speakers. The growth curve analysis revealed no significant difference between the T4 in T4.T5 and the T1 in T1.T5 in overall f0 mean (Estimate = -0.11, t = -1.27, p > .05). But, the two tones are significantly different in steepness of f0 slope (Estimate = 1.97, t = 6.89, p < .001 ***) and sharpness of the centered peak (Estimate = -0.20, t = -1.45, p < .05 *). The T4 in T4.T5 is significantly less steep in f0 slope and less sharp in the centered peak than the T1 in T1.T5, and thus the sandhied T4 in T4.T5 is non-neutralizing with its claimed surface tone T1 in T1.T5. Hence, we conclude that dissimilatory Sandhi 2 applies in a gradient way.

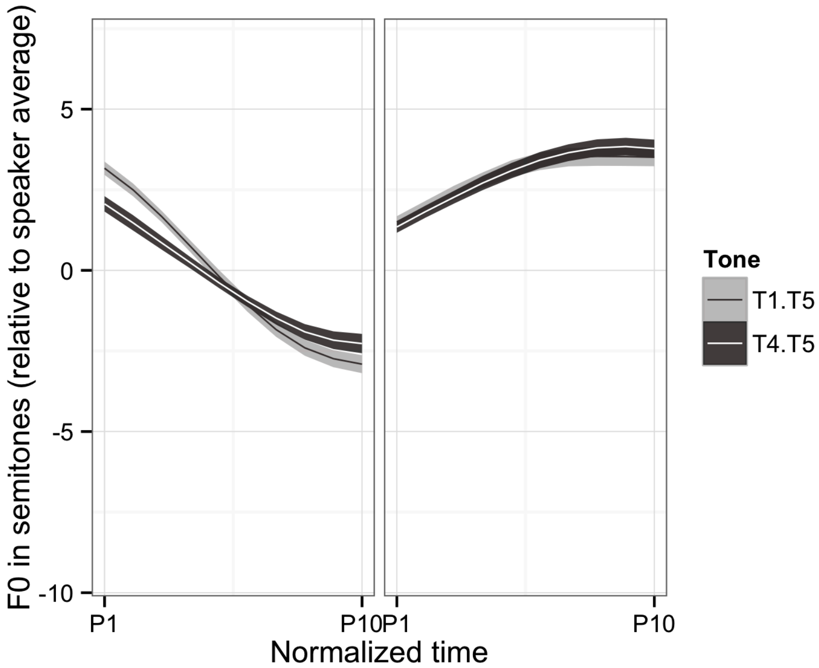

Figure 4. *Sandhied T4.T5 vs. its claimed surface T1.T5 in Nanjing Mandarin (with normalized time) by 18 speakers. The lefthand panel shows the pitch contours (in semitones relative to speaker average) of the first syllables; the righthand panel shows those of the second syllables. Central lines represent the means. Shaded areas (ribbons) stand for ±1 standard error of the mean.*

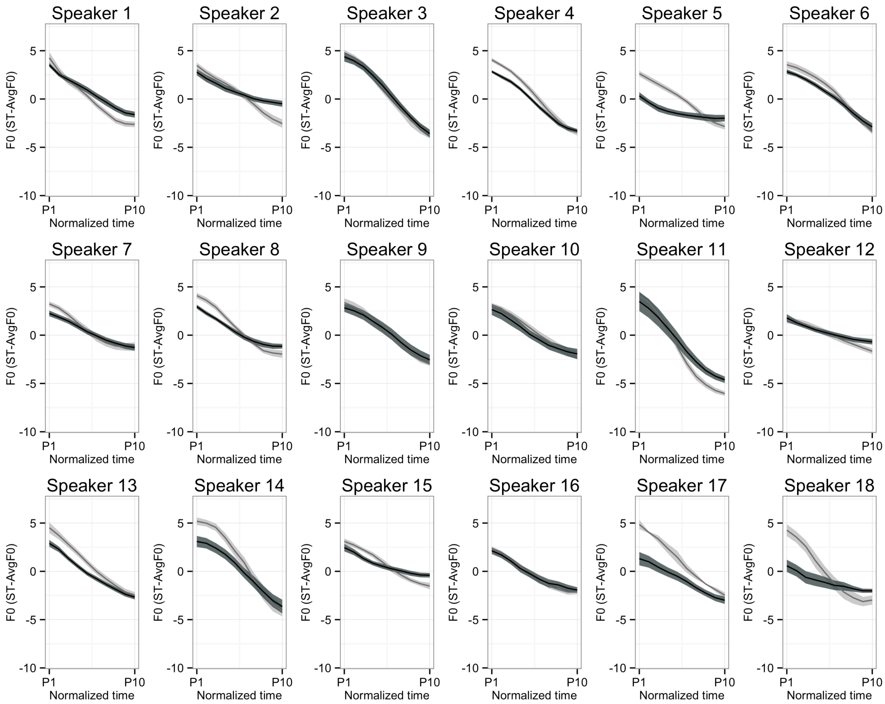

Figure 5. *T4 in T4.T5 and T1 in T1.T5 produced by different speakers (the darker shades stand for the T4 in T4.T5, the lighter shades stand for the T1 in T1.T5). ST-AvgF0: semitone relative to speaker average.*

From the best-fitting model we extracted the random effect (intercept, linear, and quadratic) estimates of the T4 in T4.T5 and the T1 in T1.T5 for each speaker (for a detailed report on the estimates, see Appendix B). The standard deviations for the intercept, linear, and quadratic differences (between the T1 in T1.T5 and the T4 in T4.T5) across the speakers were computed following the same procedures in Sandhi 1. The standard deviation for the intercept, linear, and quadratic difference (between the T1 in T1.T5 and the T4 in T4.T5) across speakers is 0.43, 1.05, and 0.20, respectively, which suggests speaker variation in overall f0 mean, f0 slope, and sharpness of the centered peak differences between the two tones.

As Figure 5 shows, four speakers (speakers 3, 9, 10 and 16) evidently applied the tone sandhi process in a categorical way; two/three speakers, i.e., speakers 5 and 18 (and 17), did not apply sandhi at all; all the other speakers seemed to apply the sandhi process in a gradient way.

2.4.2.2 Sandhi Pair 2: assimilatory Sandhi 3 vs. dissimilatory Sandhi 4

2.4.2.2.1 Categoricalness/gradience of Sandhi 3

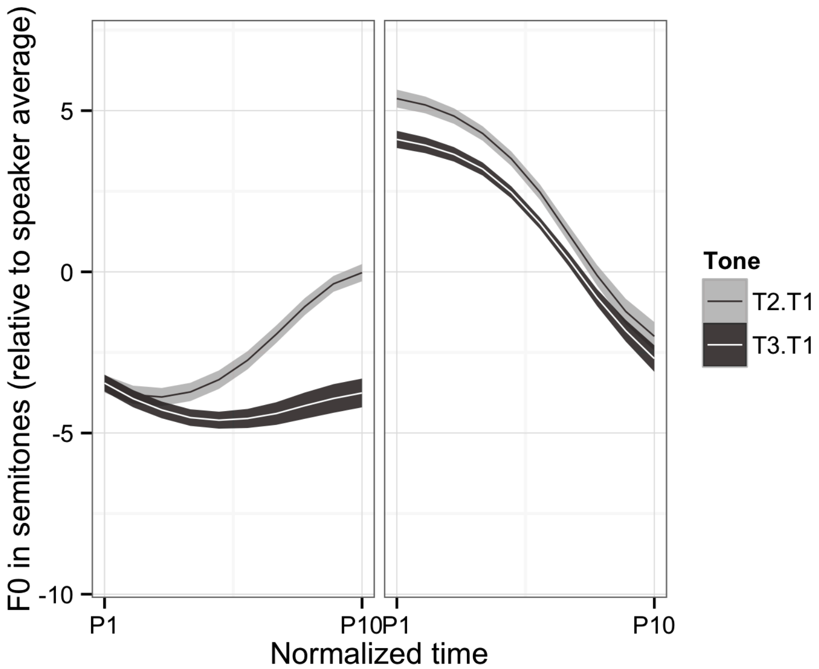

Figure 6. *Sandhied T3.T1 vs. its claimed surface T2.T1 in Nanjing Mandarin (with normalized time) by 18 speakers. The lefthand panel shows the pitch contours (in semitones relative to speaker average) of the first syllables; the righthand panel shows those of the second syllables. Central lines represent the means. Shaded areas (ribbons) stand for ±1 standard error of the mean.*

Figure 6 depicts the sandhied form of T3.T1 against its claimed target surface tone pattern T2.T1, with averaged data across 18 speakers. The growth curve analysis revealed a significant difference between the T3 in T3.T1 and the T2 in T2.T1 in overall f0 mean (Estimate = -1.55, t = -19.67, p < .001 ***), f0 slope (Estimate = -3.61, t = -9.57, p < .001 ***), but no significant difference in sharpness of the centered peak (Estimate = -0.35, t = -1.99, p > .05). Hence the sandhied T3 in T3.T1 is evidently non-neutralizing with its claimed surface tone T2 in T2.T1, suggesting that assimilatory Sandhi 3 applies in a gradient way.

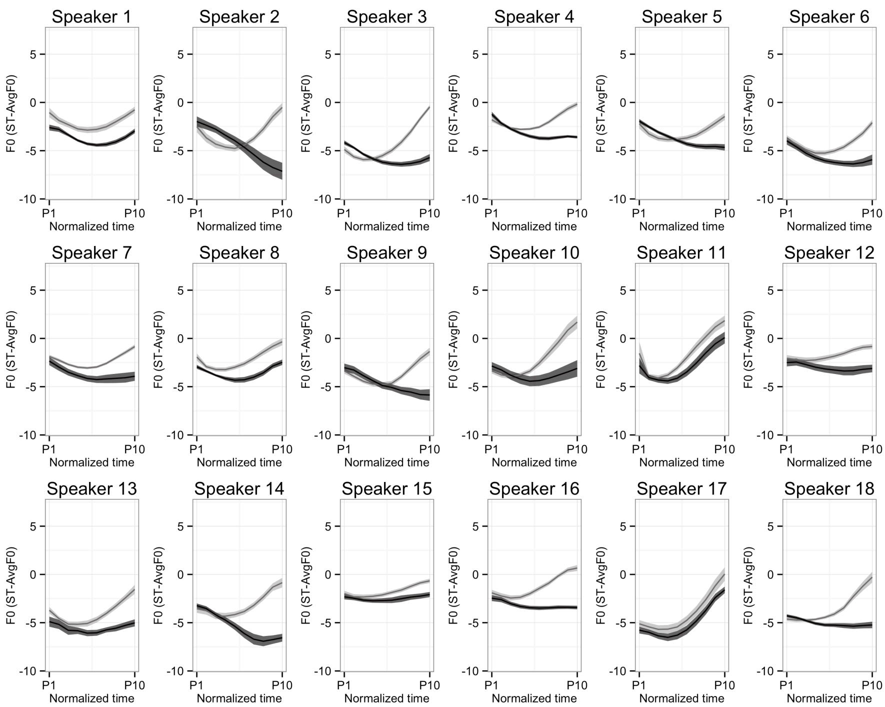

Figure 7. *T3 in T3.T1 and T2 in T2.T1 produced by different speakers (the darker shades stand for the T3 in T3.T1, the lighter shades stand for the T2 in T2.T1). ST-AvgF0: semitone relative to speaker average.*

The random effect (intercept, linear, and quadratic) estimates of the T3 in T3.T1 and the T2 in T2.T1 for each speaker were extracted from the best-fitting model (for a detailed report on the estimates, see Appendix B). The standard deviations for the intercept, linear, and quadratic differences (between the T2 in T2.T1 and the T3 in T3.T1) across the speakers were computed following the same procedures in Sandhi 1. The standard deviation for the intercept, linear, and quadratic difference (between the T2 in T2.T1 and the T3 in T3.T1) across speakers is around 0.41, 2.35, and 0.61, respectively, indicating speaker variation in overall f0 mean, f0 slope, and sharpness of the centered peak differences between the two tones. As Figure 7 demonstrates, at least three speakers evidently did not apply the sandhi process, e.g., speakers 2, 5, and 9 produced a low falling f0 contour, with no rising tendency at all; two speakers, speakers 11 and 17, evidently applied the tone sandhi process by producing a rising pitch contour resembling the T2 in T2.T1; none of the speakers applied sandhi in a categorical way.

2.4.2.2.2 Categoricalness/gradience of Sandhi 4

Figure 8 depicts the sandhied form of T3.T3 against its claimed target surface tone pattern T2.T3, averaged across 18 speakers. The growth curve analysis revealed that the T3 on syllable 1 (T3 (syl1)) in T3.T3 and the T2 in T2.T3 are not significantly different in f0 mean (Estimate = -0.13, $t = -1.72$, $p > .05$), f0 slope (Estimate = 0.16, $t = 0.66$, $p > .05$), or sharpness of the centered peak (Estimate = 0.05, $t = 0.37$, $p > .05$). Therefore there was no reason to reject the null hypothesis that the sandhied form of T3 (syl 1) in T3.T3 is neutralizing with its claimed surface tone T2 in T2.T3, which suggests that assimilatory Sandhi 4 applies in a categorical way.

The random effect (intercept, linear, and quadratic) estimates of the T3 (syl1) in T3.T3 and the T2 in T2.T3 for each speaker were extracted from the best-fitting model (for a detailed report on the estimates, see Appendix B). The standard deviations for the intercept, linear, and quadratic differences (between the T2 in T2.T3 and the T3 (syl1) in T3.T3) across the speakers were computed following the same procedures in Sandhi 1. The standard deviation for the intercept, linear, and quadratic difference (between the T2 in T2.T3 and the T3 (syl1) in T3.T3) across speakers is around 0.34, 0.73, and 0.20, respectively, which suggests speaker variation in overall f0 mean, f0 slope, and sharpness of the centered peak differences between the two tones. As Figure 9 displays, speakers 11 and 17 may have applied the tone sandhi process in a gradient way; the rest of the speakers applied sandhi in a categorical way; speaker 15 even over-applied the sandhi process, with the sandhied tone produced in a steeper rise than the T2 in T2.T3.

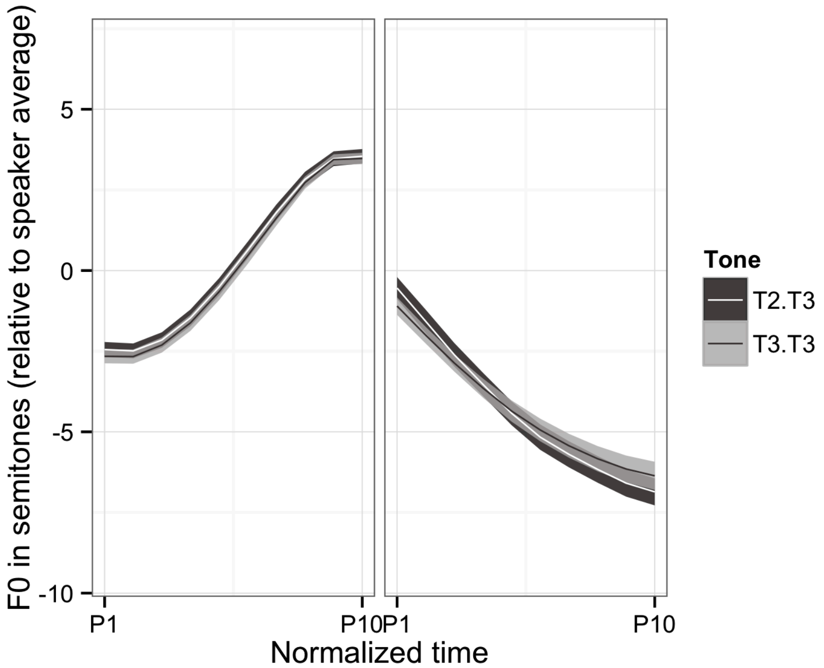

Figure 8. *Sandhied T3.T3 vs. its claimed surface T2.T3 in Nanjing Mandarin (with normalized time) by 18 speakers. The lefthand panel shows the pitch contours (in semitones relative to speaker average) of the first syllables; the righthand panel shows those of the second syllables. Central lines represent the means. Shaded areas (ribbons) stand for ±1 standard error of the mean.*

.

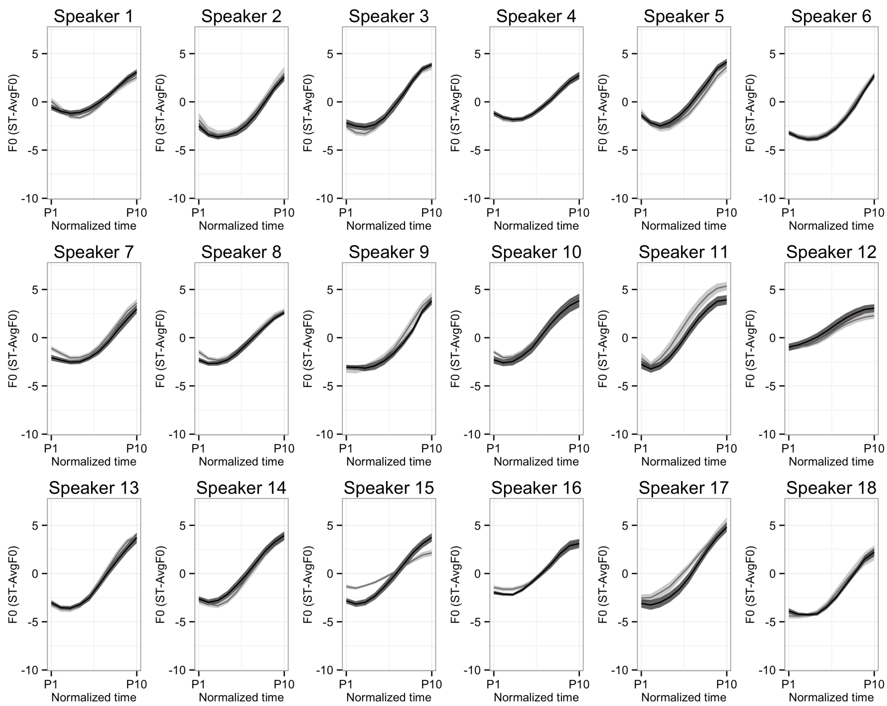

Figure 9. *T3(syl 1) in T3.T3 and T2 in T2.T3 produced by different speakers (the darker shades stand for the T3(syl 1) in T3.T3, the lighter shades stand for the T2 in T2.T3). ST-AvgF0: semitone relative to speaker average.*

Table 4 shows the speaker variation in the differences between the sandhied tone and the claimed surface tone in the four tone sandhi processes in standard deviation. From the table, we can tell that the amount of speaker variation in producing the four sandhi processes was different; the speakers applied the sandhi process most consistently in Sandhi 4, with the smallest values in two aspects of tone comparison between the sandhied tone and the claimed output tone (overall f0 mean, f0 slope); the speakers applied sandhi most inconsistently in Sandhi 3, shown by the largest values in two aspects of tone comparison between the sandhied tone and the claimed output tone (namely f0 slope and sharpness of centered peak).

Table 4. *Summary of speaker variation in the differences (in standard deviation (SD), rounded off to two decimals) between the sandhied tone and the claimed surface tone in the four sandhi processes.*

	Sandhi 1	Sandhi 2	Sandhi 3	Sandhi 4
SD of intercept differences (Overall f0 mean)	0.74	0.43	0.41	0.34
SD of linear differences (F0 slope)	0.84	1.05	2.35	0.73
SD of quadratic differences (Sharpness of centered peak)	0.18	0.20	0.61	0.20

2.5 General discussion and conclusions

The current study aimed to investigate the categoricalness/gradience of tone sandhi application in assimilatory and dissimilatory tone sandhi processes in a single language, in order to avoid interference from language-specific factors. The hypotheses were: a) assimilatory processes can be gradient, as predicted by Kiparsky; b) dissimilatory processes should always be categorical and never apply in a gradient fashion, as predicted by Ohala's "hyper-correction" theory. Hence, any findings of gradient dissimilatory processes will disfavor the "hyper-correction" view of dissimilation.

We selected two pairs of assimilatory and dissimilatory tone sandhi processes in Nanjing Mandarin: Pair 1 involving T1 vs. T4 tonal contrast and Pair 2 involving T3 vs. T2 contrast, based on previous researchers' observations. Pair 1 involves Sandhi 1: T1.T1 → T4.T1 and Sandhi 2: T4.T5 → T1.T5. Assimilatory Sandhi 1 changes an underlying high falling T1 before another T1 to a surface high level T4, while dissimilatory Sandhi 2 alters a high level T4 before a high salient T5 to a high falling T1. Pair 2 involves Sandhi 3: T3.T1 → T2.T1 and Sandhi 4: T3.T3 → T2.T3. Assimilatory Sandhi 3 changes an underlying low level/dipping T3 before a high falling T1 to a surface low rising T2, while dissimilatory Sandhi 4 alters a low level/dipping T3 before another T3 also to a low rising T2.

In Sandhi Pair 1, assimilatory Sandhi 1 was observed to be a near-categorical tone sandhi process; dissimilatory Sandhi 2 to be a gradient process. In Sandhi Pair 2, assimilatory Sandhi 3 was found to be a gradient sandhi process; dissimilatory Sandhi 4 to be a categorical sandhi process.

The observations of near-categoricalness in assimilatory Sandhi 1 and gradience in assimilatory Sandhi 3 are both congruent with Kiparsky's prediction that assimilatory processes are allowed to apply in a gradient fashion.

Though we found that dissimilatory Sandhi 4 is a categorical process, our observation of gradience in dissimilatory Sandhi 2 suggests that dissimilatory changes do not necessarily apply in a discrete way; instead, the dissimilatory changes can also occur in a gradient way, which contradicts the prediction made by the "hyper-correction" account that dissimilatory changes should always be categorical. Therefore, our results seem to lend more support to the competing "motor planning" account of the motivation for dissimilation, at least if we subscribe to the prediction made by Garrett (2015), also assumed by Jatteau and Hejná (2016), that dissimilatory processes are allowed to be gradient in their initial stages. However, we are cautious about this interpretation, because the prediction made by Garrett and Johnson (2012) was different (dissimilatory processes caused by "motor planning" tend to apply categorically, as discussed above in Section 2.1).

Phonetically gradient dissimilatory processes have also been observed in other languages/dialects in both segmental and tonal studies. For example, Jatteau and Hejná (2016) observed in Aberystwyth English (Wales) a gradient aspiration dissimilation process, which reduced (but not lost) the second aspiration feature of a C^hVC^h sequence. Similar results were found in dissimilatory aspiration processes in Halh Mongolian (Jatteau & Hejná, 2018; Svantesson & Karlsson, 2012; Svantesson, Tsendina, Karlsson, & Franzén, 2005) and in Georgian (Begus, 2016). Regarding tonal studies, Jie Zhang and Liu (2011) observed in Tianjin Mandarin several gradient dissimilatory tone sandhi processes. In addition to the aforementioned process that turns a lexical low dipping-rising T3 preceding another T3 to a tone resembling a lexical mid-rising T2 but with relatively lower overall mean pitch, they also found a low dipping-rising T3 preceding a high falling T4 reduces its rising tail and changes to a tone between the underlying T3 and a lexical mid falling T1. However, some of the studies were based on limited data, e.g., Svantesson et al. (2005) and Svantesson and Karlsson (2012) were based on only one speaker. Only Jie Zhang and Liu (2011) discussed inter-speaker variation; the rest of the studies did not make it clear whether the observed gradiency was a nature of the dissimilatory processes, or alternatively, was simply a result of large speaker variation.

In the current study, we observed inter-speaker consistency differences (see Table 4) in the application of the four tone sandhi processes, expressed in standard deviations. Among the four sandhi processes, assimilatory Sandhi 3 applies most inconsistently across different speakers; dissimilatory Sandhi 4 applies across speakers in the most consistent way. The difference in speaker consistency may be due to historical factors related to the tone sandhi processes, i.e., an ongoing sandhi process may be still on its way spreading from a small population of speakers to a larger population; whereas a historically completed sandhi rule may have already completed this spreading process, and may have already been acquired as part of phonological grammar by the majority of the native speakers. For instance, Sandhi 4 in Nanjing Mandarin: T3.T3 → T2.T3 seems to correspond to a historical sandhi pattern in Mandarin turning a *shang* (in traditional tone classification of this dialect) tone (low level) preceding another *shang* tone to a *yangping* (low rising) tone, which was documented in textbooks in the 16th century (Mei, 1977) and is probably a highly completed sandhi process. It is observed in this study to apply consistently across almost all speakers.

The findings that the four tone sandhi rules discussed in this study differ in inter-speaker consistency highlight a problem in most of previous experimental tonal studies, which only used averaged f0 means across speakers to conclude tone sandhi patterns, ignoring the variation across speakers. Yet inter-speaker consistency is an important aspect of tone sandhi production, and ignoring it may cause confusion in understanding the nature of tone sandhi processes. For example, an optional but categorical tone sandhi process can appear as a gradient process if we only look at the averaged means. We propose that inter-speaker consistency should standardly be reported in future experimental studies on tone sandhi, in order to give a full picture of tone sandhi realizations.

The current study did not control for segmental features used in the words in the assimilatory and the dissimilatory conditions. This made it difficult to compare intra-speaker variations between the conditions. We suggest that future studies adopt minimal pairs of words with the same segmental features across experimental conditions to avoid any undesired effects from segmental variation.

Chapter 3

A cross-language study on surface-to-underlying tone mapping: The role of language specificity and generality[6]

3.1 Introduction

In continuous speech, the production of one sound is typically affected by the articulatory features of its contextual sounds. When this happens, the articulated surface form of the target sound may differ from its underlying form. For instance, an underlying oral vowel may surface as nasalized due to coarticulation with an adjacent nasal consonant, as in Thai and English; an underlying coronal may surface with labial place when immediately followed by a labial, as in English, *lean bacon* can be heard as *lea*[m] *bacon* (Gaskell & Marslen-Wilson, 1996); or a labial may become a velar when immediately preceding a velar, as in Korean, /ip+ko/ 'wear and…' can surface as [ikko] (Jun, 2004; Kim-Renaud, 1974). Discrepancies between surface sound and underlying sound due to contextual change in production may pose a problem in perception for listeners (Beddor, 2009). How do listeners

[6] Preliminary results of this chapter were presented at *the 5th International Symposium on Tonal Aspects of Languages*, and published in the proceedings as X. Li, Kager, and Gu (2016).

perceive a surface sound? Are listeners able to perform the mapping between the surface sound and its underlying sound? Two general questions follow: 1) Is this surface-to-underlying mapping dependent exclusively on language-specific knowledge and is therefore restricted to native listeners? Alternatively, is this mapping assisted by language-general knowledge that is also accessible to naïve non-native listeners? 2) Is the difficulty of the surface-to-underlying mapping task influenced by different types of processes, e.g., assimilatory versus dissimilatory processes, especially for naïve non-native listeners? This chapter sets out to investigate these issues.

We hypothesize that the mapping mechanisms between the surface sound and its underlying sound are different for native listeners and naïve non-native listeners. Taking the Korean labial-to-velar assimilation rule (e.g., /ip+ko/ 'wear and…' becomes [ikko] (Jun, 2004)) as a hypothetical case, the surface form, a velar-velar sequence, contains information that is ambiguous only for native listeners (and not for non-natives), because it can underlyingly be either (a) a labial-velar sequence, in which the labial is neutralized to a velar under the influence of the contextual (following) velar, or (b) a true velar-velar sequence. For native listeners, to perform surface-to-underlying mapping means to 'undo' the labial-to-velar neutralization process and recover the underlying identity from the surface form. This mechanism should rely on native lexical knowledge or knowledge of the phonological process. By definition, naïve non-native listeners lack the language-specific phonological knowledge or lexical underlying representations to rely on in order to accomplish surface-to-underlying mapping in a foreign language.[7] What naïve non-native listeners are hypothetically doing in a surface-to-underlying mapping task is construing a link between the surface sound and the target (underlying) sound via

---

[7] Theoretically speaking, when presented with a surface-to-underlying mapping task in a foreign language, naïve non-native listeners should also be able to adopt their L1 phonological knowledge. However, in this study the key group with which we investigated the language-general effects consisted of Dutch listeners, whose non-tonal language background is maximally distant from the target tonal language (as will be discussed at the end of this section). In this way, we were able to control for any possible transfer effects, as the L1 phonological knowledge of the non-native group could not possibly influence the results in the task. An effect of L1 phonological knowledge was found in a control group of non-native tone listeners – the BJ group (as will be discussed in Section 3.3.1).

the given context, presumably through some language-general knowledge, e.g., knowledge about articulatory gestures, which listeners may have generalized from experience of articulation accumulated in their own language, or from general articulatory settings.

The mapping problem comes in two versions, depending on the type of contextual change imposed by the process. Processes of contextual change can be classified as “assimilatory” or “dissimilatory” based on whether the altered sound becomes more or less similar to its context sound. When the underlying sound receives extended articulatory features from the contextual sound, the process is defined as *assimilatory*. The examples listed above are all assimilatory in nature. When the context sound causes the target sound to change its phonetic features while these sounds underlyingly share the same phonetic features, the process is referred to as *dissimilatory*. An example is found in Southern Bavarian German, where the liquid /r/ is converted to a non-liquid /d/ before or after an adjacent liquid /l/, e.g., /tiər-lə/ ‘animal (diminutive)’ becomes [tiədlə] (Hall, 2009). In this chapter, we are mainly concerned with assimilatory and dissimilatory processes that are locally conditioned, i.e., changes triggered by an adjacent sound, although non-local contextual assimilatory and dissimilatory changes are also observed in natural languages, e.g., non-local liquid dissimilation in Latin (e.g., /sol-alis/ ‘solar’ becomes [solaris]) (Abrego-Collier, 2013)), and non-local liquid assimilation in the Bantu language Bukusu (e.g., /kar-il-a/ ‘twist’ becomes [kar-ir-a] (Odden, 1994)).

While assimilatory and dissimilatory processes are well attested in many languages, assimilatory phenomena are far more prevalent than dissimilatory phenomena across the languages of the world (e.g., Grammont, 1933; Johnson, 1973). This difference in their prevalence may be partly due to their origins. Assimilatory processes are mostly believed to be grounded in articulation. Many scholars have conceived these processes as motivated by increasing ease of articulation (e.g., Grammont, 1933; Johnson, 1973). Among the various views accounting for dissimilatory changes, the most widely accepted view was proposed by Ohala (1993), who posited that a dissimilatory sound change occurs because listeners are under pressure to preserve un underlying segment in the surface form, and misattribute a feature that is intrinsic to the surface form to coarticulation, erroneously applying corrective processes (“hyper-correction”). An alternative proposal regarding the motivation of dissimilatory processes is the "motor planning" account (Frisch, 2004; Frisch et al., 2004; Garrett, 2015; Garrett & Johnson, 2012; Grammont, 1895, 1933; Tilsen, 2007). "Motor planning" is generally viewed as the process of "constructing or retrieving motor plans that will later be executed by speaking" (Garrett & Johnson, 2012, p. 59). Under this view, the motivation for dissimilatory processes resides in difficulties on the part of the speaker regarding the planning of sequences of identical/similar sounds.

From the different origins of assimilatory and dissimilatory processes, we could hypothesize the language-general/specific nature of surface-to-underlying mapping in these processes. The motor theory (Liberman et al., 1967; Liberman & Mattingly, 1985) and the Direct Realist Theory (Fowler, 1986, 1996, 2006; Fowler & Brown, 2000) agree on the idea that acoustic events are perceived as being caused by articulatory gestures. Evidence to support this idea can be found in the literature. For instance, McGurk and MacDonald (1976) found that the perception of a speech sound is distracted by the visual information of a mouth performing the articulatory gestures of another sound (e.g., /ba/ is perceived as /da/ when a mouth articulating /ga/ is presented) (the "McGurk effect"). Moreover, D'Ausilio et al. (2009) found that magnetic stimulation of lip-related areas in the motor cortex facilitated discrimination of lip-related speech sounds (/ba/, /pa/) but not tongue-related sounds (/da/, /ta/) presented in noise, whereas stimulation of tongue-related motor cortex areas produced the reverse effect (D'Ausilio et al., 2009). The idea that articulatory ease provides the motivation for assimilatory processes would suggest that the surface-to-underlying mapping for these processes is possibly language-general, i.e., accessible for both native and naïve non-native listeners of the target language (the language-general mapping hypothesis for assimilatory processes). In contrast, for dissimilatory processes, the "hyper-correction" view implies that their surface-to-underlying sound mapping should be language-specific, i.e., only accessible for natives, and inaccessible for naïve non-native listeners, because the lexical underlying representations that this view crucially refers to are, by definition, unavailable for naïve non-native listeners (the language-specific mapping hypothesis for dissimilatory processes). The "motor planning" view, nevertheless, does not derive clear and consistent predictions regarding the language specificity/generality of the mapping for dissimilatory processes. Consequently, we will only test the yet-undisputed language-specific mapping hypothesis predicted by the "hyper-correction" approach for dissimilatory processes.

Nevertheless, experimental evidence supporting the language-general mapping hypothesis for assimilatory processes and the language-specific mapping hypothesis for dissimilatory processes is missing. A comparison between the surface-to-underlying mapping for native and non-native listeners may reveal the generality versus language experience in terms of mapping in assimilatory and dissimilatory processes. This is what we will do in this study.

This study looks into lexical tone and tone sandhi phenomena to investigate the above issue. Tone sandhi processes can also be classified as either assimilatory or dissimilatory. For example, in Zhenjiang Mandarin (Qiu, 2012), an assimilatory tone sandhi rule turns a high falling tone immediately before another high falling tone into a high level tone, with its tone offset raised to the same pitch (f0) height of its context tone onset (42.42 → 44.42) (the numbers indicate tone values on Chao (1930)'s five-level scale, with 1 representing the lowest pitch level in the speaker's pitch range, and 5 representing the highest; the dot is used to separate pitch values of the tones). The famous Tone 3 sandhi rule in Beijing Mandarin is an example of the dissimilatory tone sandhi process (Cheng, 1973; Yip, 1980, 2002): a low dipping Tone 3 immediately preceding another low dipping Tone 3 changes to a low-rising Tone 2 (214.214 → 24.214), (e.g., /maiT3 maT3 买马/ 'to buy a horse' becomes [maiT2 maT3 埋马] 'to bury a horse') The distinctive feature of lexical Mandarin Tone 3 is "low." This tone sandhi rule is often transcribed as LL.LL → LH.LL (e.g., Yip, 1980) with the altered tone offset deviating from the subsequent context tone onset. In order to subject the language-general surface-to-underlying mapping hypothesis for assimilatory processes to the strongest possible test, we intend to keep the non-native listener group as naïve as possible. By using the phenomenon of lexical tone, we will be able to test a group of listeners that is maximally remote from the native language group, namely a group of non-tonal language listeners without any previous exposure to tones who are thus guaranteed to have no articulatory experience with tone sandhi.

Native listeners are predicted to be capable of undoing the underlying-to-surface neutralization induced by a context sound in both assimilatory and dissimilatory processes on the basis of language-specific lexical or phonological knowledge. Naïve non-native listeners, however, are predicted to construe the link between a surface and a target (underlying) sound only in assimilatory processes, via language-general knowledge about articulatory ease.

The literature on surface-to-underlying mapping by native and naïve non-native listeners does not include all logical scenarios, and studies on tonal mapping are particularly few. Native listeners are widely found to be able to map a surface sound to its underlying counterpart by perceptually undoing any neutralization effects attributed to context and successfully recovering the underlying sound. For instance, a reversed version of the coronal place assimilation process occurs in English, German and Dutch, in which native listeners undo the neutralization of place due to the context sound and perceive a word-final labial, when followed by a word-initial labial, as a coronal (German: Coenen et al., 2001; English: Darcy et al., 2009; Gaskell & Marslen-Wilson, 1996, 1998; Gaskell & Snoeren, 2008; Gow, 2003; Dutch: Mitterer, 2003), e.g., *lean bacon* is successfully recovered from *lea*[m] *bacon* (Gaskell & Marslen-Wilson, 1996). Examples can also be seen in vowel nasalization in English (Beddor & Krakow, 1999; Lahiri & Marslen-Wilson, 1991), voicing assimilation in Hungarian (Gow & Im, 2004), French (Darcy et al., 2009), and German (Coenen et al., 2001), and labial-to-velar assimilation in Korean (Mitterer et al., 2013). Among these studies, Mitterer et al. (2013) found that the most assimilated sounds, which were most ambiguous to native listeners, lead to more successful surface-to-underlying mappings for the listeners, compared to the less assimilated sounds, in a categorization experiment. This observation suggested that the native surface-to-underlying mapping in assimilatory processes probably does not rely on the presence of residual phonetic cues for the underlying sound in the surface pattern, but rather relies on the lexical knowledge or phonological knowledge acquired in one's native language. Some studies engineered experimental conditions in a way that lexical knowledge could not be made use of in the mapping task. For instance, Gaskell and Marslen-Wilson (1998) adopted non-lexical words such as *preight* in a phoneme monitoring experiment, and found that native English participants heard /t/ (a part of /preɪt/ (*preight*)) more often in /preɪp beərə/ (*preight bearer*), where the context supported the place assimilation, rather than in /preɪk beərə/, where there was no clue about the place assimilation.

The results indicated that the native mapping process which recovers the underlying form from the surface form operates on non-words as well, suggesting that the native mapping is not based entirely on access to lexical knowledge. Tonal studies added to the evidence that native listeners use the variation due to assimilatory processes in a phonological way to recover the underlying tone identity. For example, Xu (1994) presented naturally produced Mandarin trisyllabic nonwords to native Beijing Mandarin listeners and asked them to identify the tonal identity of the coarticulated rising/falling tone on the middle syllable. The native listeners recovered the underlying tone identity of the target tone at an overall high accuracy, despite the significant distortion on the underlying tone in many cases, showing that the mapping did not rely on phonetic residues. This study also used nonsense strings to make sure that the task would not remind listeners of real words, thus providing evidence that abstract phonological knowledge was used in the mapping.

Regarding dissimilatory processes, few segmental studies have explored native listeners' capability of mapping the surface form to the underlying representation in them. Some tonal studies, such as Peng (2000), A. Chen and Kager (2011), A. Chen et al. (2015) and Nixon, Chen, and Schiller (2015) investigating the Mandarin Tone 3 sandhi rule (T3.T3 → T2.T3) considered dissimilatory in nature as discussed earlier (e.g., Cheng, 1973; Yip, 1980, 2002), offer some evidence in favor of the sandhied tone being perceived by native listeners as an underlying Tone 3. Peng (2000) tested if native Mandarin listeners categorized the derived Tone 2 (sandhied Tone 3) of the Mandarin Tone 3 sandhi rule followed directly by the contextual Tone 3 as belonging to the underlying Tone 3 category in a Concept Formation experiment (this paradigm will be discussed in detail later), and found that the listeners immediately and consistently (> 80%; chance level: 50%) categorized the surface Tone 2 as belonging to this category. Moreover, Nixon et al. (2015) observed in a picture naming task that native Beijing Mandarin listeners were faster in naming pictures of disyllabic words containing a lexical Tone 3 with superimposed distractors of disyllabic words of sandhied Tone 3, which can be viewed by us as additional evidence of the sandhied Tone 3 mapped onto the underlying T3. These two studies adopted test words whose surface disyllabic words do not exist in the Mandarin lexicon whereas their underlying disyllabic words do, e.g., in Peng (2000), test words were like /y35san214/: the surface word /y35san214/

is a gap in the Mandarin lexicon, whereas the underlying word /y214san214 雨伞/ means ‘umbrella.’ Therefore the design of the stimuli invited lexical knowledge to play a major role in the task, and did not convince us that the observed surface-to-underlying mapping was due to abstract phonological knowledge. A. Chen and Kager (2011) and A. Chen et al. (2015) designed their studies in a way as to limit the role of lexical knowledge. A. Chen and Kager (2011) found that Mandarin Tone 3 and Tone 2 were discriminated by native listeners more accurately when Tone 3 preceded Tone 2 than when Tone 2 preceded Tone 3, indicating that the Tone 2 in the Tone 2 Tone 3 sequence might be perceived by the native listeners as a sandhied tone with un underlying identity of Tone 3. In this study all tones were realized on the syllable /ma/. Though /maT3 马/ ‘horse’ is an existing word in Mandarin as is /maT2 麻/ ‘hemp,’ the underlying disyllabic /maT3 maT3/ does not exist in the Mandarin lexicon. In a following disyllabic tone discrimination study, A. Chen et al. (2015) found that native Mandarin listeners confused the disyllabic sequences /Tone 2 Tone 3/ and /Tone 3 Tone 3/ more than the sequences /Tone 3 Tone 2/ and /Tone 3 Tone 3/, although the words they used bearing /Tone 3 Tone 3/ are gaps in the Mandarin lexicon. The results also suggest that the native listeners may have mapped the Tone 2 in the /Tone 2 Tone 3/ string to its underlying identity Tone 3. Though these two studies adopted Mandarin characters, they avoided Peng (2000)’s use of existing disyllabic words and non-existing disyllabic words as the underlying words and the surface words, respectively. Hence the lexicon could not have been the reason that directly lead the listeners to map the surface words onto the underlying words in this study; abstract phonological knowledge must have played a greater role.

With respect to naïve non-native listeners, some empirical evidence can be found in consonantal studies regarding their access to the mapping between surface and target (underlying) sounds in some assimilatory processes. For instance, Gow and Im (2004) observed that naïve English listeners were facilitated in monitoring an underlying sound in a Hungarian voicing assimilation process in its surface form in a ‘viable context’ which licensed the assimilatory change, as compared to an ‘unviable context’ which did not support the change. Darcy et al. (2009) tested naïve French listeners on an English place assimilation process, as well as naïve English listeners on a French voicing assimilation process, and observed that both groups of listeners detected the underlying target word more successfully in viable contexts than in unviable contexts. This effect was small but statistically significant. These observations are evidence in favor of the language-general mapping hypothesis for assimilatory processes.

The findings on naïve non-native listeners' perception of assimilatory processes raise the question as to whether naïve non-native listeners could also acoustically relate the surface and target (underlying) forms in dissimilatory processes. However, to our knowledge, no previous study has investigated this issue. Hence, whether the surface-to-underlying mapping for dissimilatory processes requires language-specific knowledge, as the "hyper-correction" theory implies, remains unknown.

The current study investigates the perception of sandhied tones targeting their underlying tones in assimilatory and dissimilatory tone sandhi processes by native and naïve non-native listeners, and aims to answer the following main research questions: 1) Will native tonal language listeners be the only group to successfully perform surface-to-underlying tone mapping in the target tone sandhi processes, or will a non-tonal language group, naïve to tone grammar, also be able to perform the mapping based on language-general articulatory knowledge? 2) Will assimilatory tone sandhi processes (compared to dissimilatory processes) lead to facilitated surface-to-underlying mapping for naïve non-tonal language listeners?

A pair of assimilatory and dissimilatory tone sandhi rules are found in Nanjing Mandarin, a dialect in the Hongchao dialectal area of Jianghuai Mandarin in China, in X. Li and Kager (2018a) (see Chapter 2). Nanjing Mandarin 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 tone sandhi rules, documented in earlier studies (Jiangsusheng Difangzhi Bianzuan Weiyuanhui (Editorial Committee of Jiangsu Province Chorography), 1998; Liu, 1995; Song, 2006). We recorded 18 native speakers of Nanjing Mandarin aged 18 to 30 years old in a disyllabic tone sandhi elicitation task to produce natural disyllabic tone sandhi patterns, and observed slightly different results from previous studies. We found two tone sandhi rules differing in their assimilatory/dissimilatory nature. Figures 1 and 2 depict the averaged pitch contours of the two sandhi rules across the 18 native speakers. 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 an assimilatory tone sandhi process; Sandhi 2 is a dissimilatory tone sandhi 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 produced a steeper fall much resembling a T1 as the sandhied tone in Sandhi 2, compared to the averaged contour in the middle panel of Figure 2. Sandhi 1 changes an underlying T1 to a surface T4, and Sandhi 2 changes an underlying T4 to a surface T1. The underlying-to-surface tonal changes in the two sandhi rules are opposite in direction, yet both sandhi rules involve a Nanjing Mandarin T1 vs. T4 contrast with equal distance between the underlying tone and the surface tone in terms of tonal categories. The two sandhi rules are used as the base for creating stimuli for this study.

- Sandhi 1: T1.T1 → T4.T1

  (e.g., *xian*T1 *hua*T1 鲜花 'fresh flowers' → [xianT4 huaT1 献花] 'to present flowers')

- Sandhi 2: T4.T5 → T1.T5

  (e.g., *ban*T4 *zhuo*T5 半桌 'half a table' → [banT1 zhuoT5 搬桌] 'to move a table')

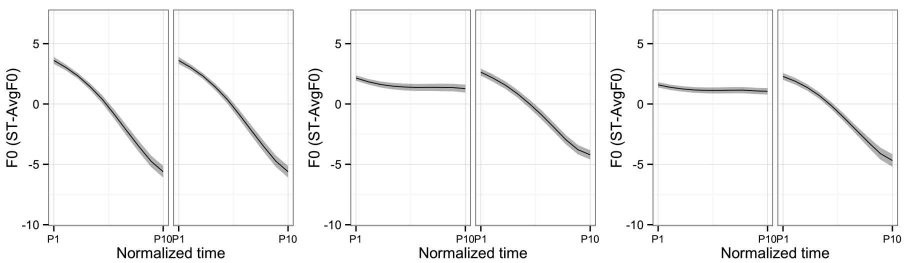

Figure 1: *Lexical T1+T1 (left), sandhied T1+T1 (middle) and T4+T1 (right) in Nanjing Mandarin (averaged across 18 speakers). Solid lines indicate mean f0 (in semitones relative to speaker average); gray ribbons stand for ±1 standard error of the mean (also in Figure 2).*

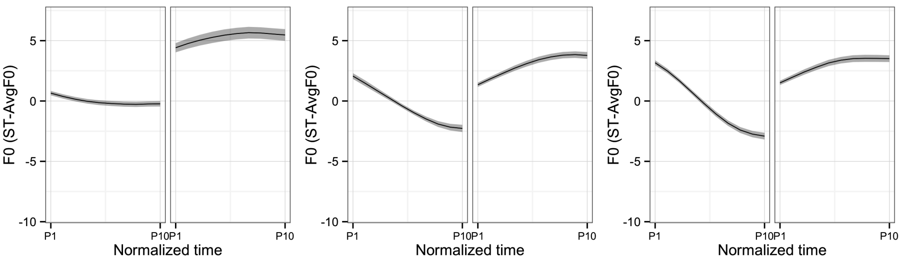

Figure 2: *Lexical T4+T5 (left), sandhied T4+T5 (middle) and T1+T5 (right) in Nanjing Mandarin (averaged across 18 speakers).*

In order to investigate the mapping between Nanjing Mandarin sandhied tones (surface tones) and underlying tones by native and naïve non-native listeners, the Concept Formation paradigm (Jaeger, 1986) 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. Hence it can be used to test both native and naïve non-native listeners. 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. Previous studies showed that it was effective for non-tonal language listeners to learn tone categories through auditory training (e.g., Francis et al., 2008). In this study, a test was given which includes (a) target tokens, (b) non-target tokens, and (c) test tokens that 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 for bringing out the latter category. Peng (2000) successfully used the Concept Formation paradigm to test categorization of the sandhied tone of the Mandarin Tone 3 sandhi rule by native Mandarin listeners. In her study, the ambiguous test tokens always contained the sandhied tone. This tone is ambiguous in that it matches both the underlying tone category and the surface tone category for native listeners. Due to its successful implementation in Peng's study, we also adopt the Concept Formation paradigm to test the categorization of sandhied tones in our study. This paradigm allows us to use either the underlying tone or the surface tone as the target category. We are mainly interested in the mapping between the sandhied tone and the underlying tone. Yet by including the surface tone as the target category, we will be able to a) for native listeners, confirm whether the sandhied tone is mapped only onto the underlying tone category or can be mapped onto the surface tone category as well, and b) for naïve non-native listeners, observe whether they can successfully identify the surface tone as the target category and perform the mapping at surface level, providing evidence of their capability to perform the tone mapping task.

It is expected that native Nanjing Mandarin listeners, when trained with the underlying tone as the target tone category, will consistently categorize the ambiguous tokens (sandhied tone) so as to match the underlying tone; they are expected to categorize the ambiguous tokens to match the surface tone when trained with the surface tone as the target category. No difference is expected for them between the underlying tone training condition and the surface tone training condition. Naïve non-tonal language listeners who have no tonal categories in their native phonology may create a new category to perceive the target tone. When the target category is the surface tone, they are predicted to perform the mapping between the sandhied tone and the target tone as the two tones are acoustically identical. However, according to our hypothesis, when the target category is the underlying tone, they are predicted to construe the link between the sandhied tone and the target category more easily in the assimilatory condition than in the dissimilatory condition, since assimilatory processes rely on language-general mapping while dissimilatory processes depend on language-specific mapping.

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 native Nanjing Mandarin group (henceforth, NJ), a non-tonal language Dutch group (henceforth, NL), and a Beijing Mandarin group (henceforth, BJ) as a non-native tonal language control group. By including the BJ group, we shall be able to a) single out the influence from native phonological knowledge on surface-to-underlying mapping by comparing the two tonal-language groups; and

b) observe if any failure in surface-to-underlying mapping by the NL group is simply due to an incapability of performing a tone task by comparing the two non-native language groups.

As shown in the Concept Formation paradigm, the underlying tone is implemented in the current study as a phonetic tone presented to the listeners. One might argue that for naïve non-native listeners, the assumed mapping between the sandhied tone and the given underlying tone is purely mapping at surface level because the listeners lack underlying forms and will hence only take a given underlying tone as its surface value. This view, nonetheless, would not change our hypotheses of language-general mapping for assimilatory processes and language-specific mapping for dissimilatory processes because the naïve non-native listeners, though possibly taking the underlying tone as its surface value, are still confronted with the task of mapping between this tone and a phonetically variant - the sandhied tone. Undoing this type of form variation should involve the same perceptual knowledge on the part of the listener as surface-to-underlying mapping. In this study, we refer to this mapping process by naïve non-native listeners as surface-to-underlying mapping consitent with the native group.

3.2 Method

As mentioned in Section 3.1, the paradigm of this study – Concept Formation – is an experiment consisting of a training session for participants to learn a category (for the NL group) or make use of an existing category (fo the NJ group), followed by a test session on an ambiguous category.

3.2.1 Stimuli

In the training session of the Concept Formation experiment, the target tokens were Nanjing disyllabic words beginning with the target tone followed by a variety of tones; non-target tokens were Nanjing Mandarin disyllabic words beginning with other tones except the target tone. No tone sandhi was involved in the training session. The number of target tokens and non-target tokens were kept equal (24)[8] (for the tone sequences used in the training session, see Appendix C). Differences in frequency would possibly lead listeners to consistently choose the more frequent token (target or non-target) as a safer strategy to attain a high correct ratio when being tested. Stimuli in the training 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 (for the tone sequences used in the test session, see Appendix C). The target and non-target tokens in the test session were new, in case some listeners simply remembered the answers directly from the learning session, but constructed in the same way as those in the training session. Again, no tone 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 is ambiguous to native listeners and matches both the target tone (underlying tone/surface tone, according to experimental condition) and its sandhi-related tone (surface tone/underlying tone).

Two sets of test tokens were created separately for a real-word session and a non-word session. The real words and the non-words 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 disyllabic word pair[9] 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

[8] In the training session, the target tokens contained 3 tone sequences which proved to preserve the target tone in disyllabic sequences (from a tone sandhi elicitation task we conducted) and included 8 tokens for each tone sequence (3 * 8 = 24 tokens). For a balanced design, the non-target tokens maintained the tones on the second syllables as in the target tokens. Depending on the presence of 8 or 9 of these tone sequences elicited a distinct tone from the target tone in the first syllable, the non-target tokens contained either 1) 8 tone sequences * 3 tokens = 24 tokens, or 2) 9 tone sequences * 3 tokens – 3 tokens = 24 tokens.

[9] The word pairs were decided by three native Nanjing Mandarin speakers.

underlying-surface pair were selected. For example, the words *xian*T1 *hua*T1 鲜花 ‘fresh flower’ and *xian*T4 *hua*T1 献花 ‘to present flowers to someone’ 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[10] was controlled by a survey of subjective rating on whether the non-words resemble words in real life. The disyllabic words with similar word-likeness ratings in the underlying-surface pair were carefully selected. For example, the words *shu*T1 *jiang*T1 and *shu*T4 *jiang*T1 are both meaningless and distant from real words in Nanjing Mandarin. Target tokens and non-target tokens in training and test sessions were also created separately for the real-word session and the non-word session, in order to match the real-word/non-word category of the test tokens (for the full list of the words used in the training session and the test session, see Appendix D).

All the tokens were spoken naturally by a 25-year-old female native speaker of Nanjing Mandarin, whose productions feature a steeper fall resembling a lexical T1 as the sandhied tone in Nanjing Sandhi 2 (as mentioned when introducing Sandhi 1 and Sandhi 2 in Section 3.1). The target tokens and non-target tokens were produced without tone sandhi. The test tokens were produced in sandhi condition. All the tokens were examined in Praat (Boersma & Weenink, 2009) and the most sandhied ones of Sandhi 1 (T1.T1 → T4.T1) and Sandhi 2 (T4.T5 → T1.T5) were selected as the auditory stimuli in the experiment.

3.2.2 Participants

For each language group (NL/BJ/NJ), 80 participants with self-reported normal hearing were recruited. All of them were aged from 18 to 30, consistent with the age span used in the previous tone sandhi elicitation task and subjective familiarity/wordlikeness-rating task in Nanjing Mandarin. None of the NJ listeners had taken part in the previous experiment, and none of the BJ listeners had been exposed to the Nanjing Mandarin tones or tone sandhi before. Listeners in each language group were randomly assigned to Sandhi 1 the assimilatory tone

[10] The word pairs were decided by the same three native Nanjing Mandarin speakers.

sandhi rule and Sandhi 2 the dissimilatory tone sandhi rule, 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 completed both the real-word session and the non-word session in one experiment. The order of the two sessions was counterbalanced between participants.

3.2.3 Procedure

In the training session, listeners were instructed to identify the melodic property which was a feature of all the target tokens they would hear and none of the non-target tokens (for the full instructions for both training and test, see Appendix E). 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 auditory feedback was in Dutch for NL listeners, in Beijing Mandarin for BJ listeners, and in Nanjing Mandarin 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[11]. Missed trials were repeated at the end of the trials, so that responses were collected for every trial.

The learning session automatically ended[12] 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 tone category[13]. Then, after a page of instructions

---

[11] For the right-handed listeners, the right shift key was set as the "target" button; for the left-handed listeners, the left shift key was the "target" button.

[12] The number of trials needed to reach this point was different for every participant. Nevertheless, the learning session was generally completed within 20 minutes.

[13] The criterion for having learned the tone category was settled according to the results of a pilot experiment with 5 NL listeners. Initially, the number of the trials in the criterion in the pilot was 15. Two of the NL listeners failed to pass the training session within 35 minutes, and reported depressed mood and fatigue. When the number was set to 13, the NL listeners showed a learning pattern: they made many mistakes in the beginning, and gradually made fewer mistakes until finally passing the criterion, generally within 20 minutes. We reasoned that the task of performing 13 randomized tone-relevant trials with varying segmental shapes in a row with 2 or fewer errors is not likely to be accomplished

(around 15 seconds reading time), they proceeded to the test session. In the test session no auditory feedback was given.

The experiment was conducted via a ZEP (Veenker, 2017) program. Both the detection (target/non-target) and reaction time for each test token were recorded. Detection rate in each condition was used as the main measure, because reaction times were calculated based on a different number of "yes" responses in the different conditions, and thus may not lead to a fair comparison between the conditions. In this chapter only the detection data are presented. A higher detection rate would suggest a better surface-to-underlying mapping.

3.3 Results and discussion

Table 1. *Model comparisons.* $\Delta\chi^2$*: change of chi-square;* Δdf*: change of degreees of freedom.*

Models	Model pairs	Comparison $\Delta\chi^2$	Δdf	p
Model 0 (base model) with 1| PARTICIPANT and 1| ITEM only				
Model 1 (+ LANGUAGE GROUP)	0 vs. 1	12.46	2	< .01 **
Model 2 (+ SANDHI)	1 vs. 2	0.52	1	.473
Model 3 (+ LANGUAGE GROUP : SANDHI)	1 vs. 3	1.98	3	.576
Model 4 (+ TRAINING CONDITION)	1 vs. 4	52.53	1	< .001 ***
Model 5 (+ LANGUAGE GROUP : TRAINING CONDITION)	4 vs. 5	182.47	2	< .001 ***

by simply remembering the tokens or by luck; instead it should rely on a newly-formed tone category.

Table 2. *Estimated parameters of fixed and random effects in Model 5.*

Fixed effects	Estimate	Standard error	z	p
Intercept (BJ Surface)	2.96	0.21	14.03	< .001 ***
NJ Surface	-1.81	0.28	-6.52	< .001 ***
NL Surface	-2.70	0.28	-9.81	< .001 ***
BJ Underlying	-5.31	0.30	-17.77	< .001 ***
NJ: Underlying[14]	5.29	0.39	13.44	< .001 ***
NL: Underlying	5.41	0.39	13.88	< .001 ***
Random effects	**Variance**	**Standard deviation**	**p**	
1| PARTICIPANT	1.12	1.06	< .001 ***	
1| ITEM	0.00	0.06	.43	

A generalized linear mixed-effects model was constructed to analyze the detection data in R (R Core Team, 2014) using the *glmer* function in the *lme4* package (Bates et al., 2014). Table 1 shows how the model was constructed. The dependent variable in the analysis was the detection value (yes/no). PARTICIPANT and ITEM were included as random effects. Fixed effects of LANGUAGE GROUP (BJ/NL/NJ), TRAINING CONDITION (underlying tone/surface tone) and the interaction between them were added to the model step by step. Log-likelihood model fit comparisons were made at each step in order to evaluate the significance of the added effects. The above three effects all proved to significantly improve the fit of the model. However, adding SANDHI (Sandhi 1/Sandhi 2) or any interaction between SANDHI and LANGUAGE

[14] Note that "NJ: Underlying" and "NL: Underlying" are interaction terms.

GROUP/TRAINING CONDITION did not significantly improve the model fit. Thus, the NL group, the BJ group, and the NJ group all failed to show a significant difference between the sandhi conditions. Table 2 shows the fixed and random effects of the model (Model 5 in Table 1) that best fits the detection data.

To allow for a post-hoc test in order to examine the difference between training conditions for each language group, as well as for a more intuitive interpretation of the estimated parameters of fixd and random effects in the table, we created in the detection data a new variable LANGUAGE GROUP & TRAINING CONDITION which by itself incorporated LANGUAGE GROUP, TRAINING CONDITION, and the interaction between them. Model 5b was constructed with this new variable as the fixed effect while keeping the rest of the model the same as Model 5. Table 3 shows the fixed and random effects of Model 5b.

Table 3. *Estimated parameters of fixed and random effects in Model 5b.*

Fixed effects	Estimate	Standard error	z	p
Intercept (BJ Surface)	2.96	0.21	14.03	< .001 ***
NJ Surface	-1.81	0.28	-6.52	< .001 ***
NL Surface	-2.70	0.28	-9.81	< .001 ***
BJ Underlying	-5.31	0.30	-17.77	< .001 ***
NJ Underlying	-1.82	0.28	-6.57	< .001 ***
NL Underlying	-2.59	0.28	-9.42	< .001 ***
Random effects	**Variance**	**Standard deviation**	**p**	
1| PARTICIPANT	1.12	1.06	< .001 ***	
1| ITEM	0.00	0.06	.43	

A Tukey post-hoc test was conducted on Model 5b to evaluate the difference between the training conditions for each language group, by using the *glht* function in the *Multcomp* package (Hothorn, Bretz, & Westfall, 2008) in R (for the full comparison results, see Appendix F). The results show that 1) the BJ group demonstrated a significant lower detection in the underlying training conditions than in the surface training condition (Estimate = -5.31, z = -17.77, p < .001 ***); 2) the NJ group did not show a significant difference between the two training conditions (Estimate = -0.11, z = -0.05, p > .05); 3) the NL group also did not show a significant difference in the mapping between the two training conditions (Estimate = 0.11, z = 0.43, p > .05).

An item analysis suggested that no specific items caused the NJ participants to fail in 100% overall mapping.

In the following sections, we will first discuss and explain the statistical results obtained for each language group. The control (BJ) group will be discussed first, because the results of this group are most explicit, and provide a base aginst which the NL group and the NJ group can be compared.

Figures 3-5 depict the probability of mapping the sandhied tone to the underlying/surface tone across conditions in each language group. From left to right, the four bars in Figures 3-5 represent 1) mapping the sandhied tone in Sandhi 1 (T4) to the underlying tone in Sandhi 1 (T1), 2) mapping the sandhied tone in Sandhi 1 (T4) to the surface tone in Sandhi 1 (T4); 3) mapping the sandhied tone in Sandhi 2 (T1) to the underlying tone in Sandhi 2 (T4), and 4) mapping the sandhied tone in Sandhi 2 (T1) to the surface tone in Sandhi 2 (T1). Mapping in the two sandhi conditions is combined in Figure 6.

3.3.1 BJ group

The results from the BJ group demonstrated a remarkable difference between the two training conditions (see Figure 3). A ceiling effect was observed in the mapping between the sandhied tone and the surface tone. Conversely, in the underlying-tone condition, most participants failed to accomplish the mapping between the sandhied tone and the underlying tone, although a few participants performed such a mapping, as Figure 6 illustrates.

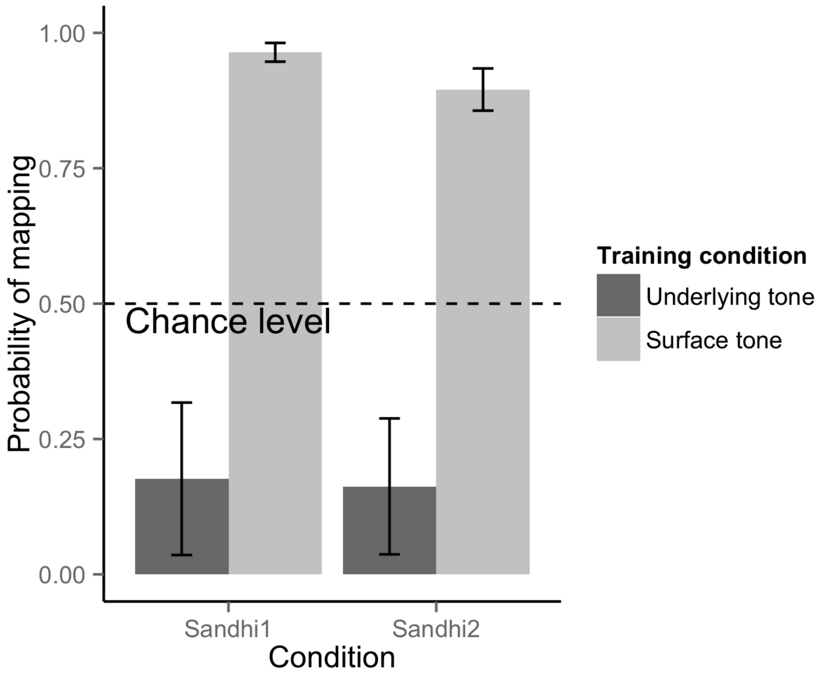

Figure 3: *Probability of mapping across conditions in the BJ group. Error bars stand for ±1 standard error of mean).*

The BJ group's failure in the mapping from the sandhied tone to the underlying tone can be explained by their native tonal grammar. The BJ group has no access to the underlying level of Nanjing Mandarin sandhied tones, because Sandhi 1 and Sandhi 2 in Nanjing Mandarin are new phonological rules to them, and they lack the necessary mapping that would allow them to construct the surface pattern of a Nanjing Mandarin sandhied tone based on its underlying properties. Instead, they take the Nanjing Mandarin underlying tone at face value and assimilate the acoustically distinct Nanjing Mandarin sandhied tone and underlying tone into distinct Beijing Mandarin tonal categories. This pattern is explained by the Perceptual Assimilation Model (PAM) by Best (Best, 1993, 1994, 1995).

According to the PAM, listeners perceptually assimilate non-native sounds to their native sound categories when encountering them; cross-language categorization is determined by the degree to which non-native sounds can be assimilated to native category representations, and similar sounds will be more easily assimilated than distant ones. Segmental studies have shown evidence in support of the PAM model in processing similar/less-similar sounds. For instance, Gilichinskaya and Strange (2010) found that American English vowels /i:/, /u:/ and /ɑ:/ are acoustically similar to Russian vowels /i/, /u/ and /a/; accordingly, naïve Russian listeners categorized the three American English vowels most consistently to the three corresponding Russian counterparts, whereas categorization of the remaining American English vowels to Russian vowels was less consistent. Tonal studies added to the evidence supporting the PAM theory, e.g., Francis et al. (2008) observed that naïve Mandarin Chinese listeners who had no prior exposure to Cantonese categorized the Cantonese high level (55) and high rising (25) tones, which have counterparts in Mandarin, accurately more than 90% of the time; whereas they accurately categorized the non-matched low level (22) tone, which lack a counterpart in Mandarin, only 22% of the time, not significantly different from the chance level of 16.67%. 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 (Peng, 1996; Speer et al., 1989; Wang & Li, 1967). Taking Sandhi 1 in Nanjing Mandarin as an example, the sandhied tone in Nanjing Mandarin Sandhi 1 (Nanjing T4) is acoustically similar to the Beijing Mandarin T1, hence it is assimilated into the Beijing T1 category; the underlying tone in Nanjing Mandarin Sandhi 1 (Nanjing T1) is acoustically similar to the Beijing Mandarin T4, hence it is assimilated into the Beijing T4 category. In sum, perceptual assimilation to native tone categories has presumably led to the Beijing listeners' consistent failure in the mapping between the sandhied tone and the underlying tone.

The BJ listeners' successful mapping from the sandhied tone to the surface tone is explained by their native tonal grammar as well. In our study, the sandhied tone and the surface tone were both Nanjing Mandarin T4 in the Sandhi 1 condition, and were both Nanjing Mandarin T1 in the Sandhi 2 condition. Nanjing Mandarin T4 is acoustically similar to Beijing Mandarin T1, hence it is assimilated into the

Beijing T1 category; and Nanjing Mandarin T1 is acoustically similar to Beijing Mandarin T4, hence it is assimilated into the Beijing T4 category. Consequently, the BJ listeners performed the mapping between the sandhied tone and the surface tone highly successfully.

No effect of sandhi condition (assimilatory Sandhi 1/dissimilatory Sandhi 2) was found in the BJ group, probably because the robust lexical tone representations allowed this group of listeners to focus exclusively on the tone on the first syllable while neglecting the tone on the second syllable as a context in the perception task.

3.3.2 NL group

Results from the NL group always fluctuated around chance level of 50% (see Figure 4). Also, this group did not demonstrate a better mapping in the surface-tone condition, where the training tone and test tone were acoustically identical rendering the task less complex, as compared to the underlying tone condition. This observation is in sharp contrast to the BJ listeners, who apparently demonstrated the adoption of their robust lexical tone categories in the mapping task. It suggests that the NL listeners probably failed to apply the newly-learned tone category in the test phase. We believe that the Dutch listeners temporally grouped the target items into a category or a quasi-category by the end of the training session, due to the fact that they passed the criterion of doing thirteen trials in a row correctly with two or fewer mistakes in the training session, which was unlikely to be accomplished by luck or by simply remembering the tokens. However, these categories/quasi-categories corresponding to the Nanjing Mandarin T1/T4 tones were weak and unstable. As such they were likely erased during the time elapsed (around 15 seconds) between the training and test when reading instructions, and hence did not assist the listeners to perform the mapping between the sandhied tone and the surface tone. A comparable study by Francis et al. (2008) found that following auditory training to recognize tones with feedback, naïve English listeners showed significant improvement in identifying Cantonese lexical tones. Though the task in their training was more challenging (identifying 6 lexical tones compared to learning 1 lexical tone for each group in the current study), their training lasted over the course of more than two weeks, which is long enough for the listeners to form more reliable categories. A comparison between the results of their study and the current study shows that in order to perform a tone learning task, naïve non-tonal language listeners needed to be trained for a longer period of time.

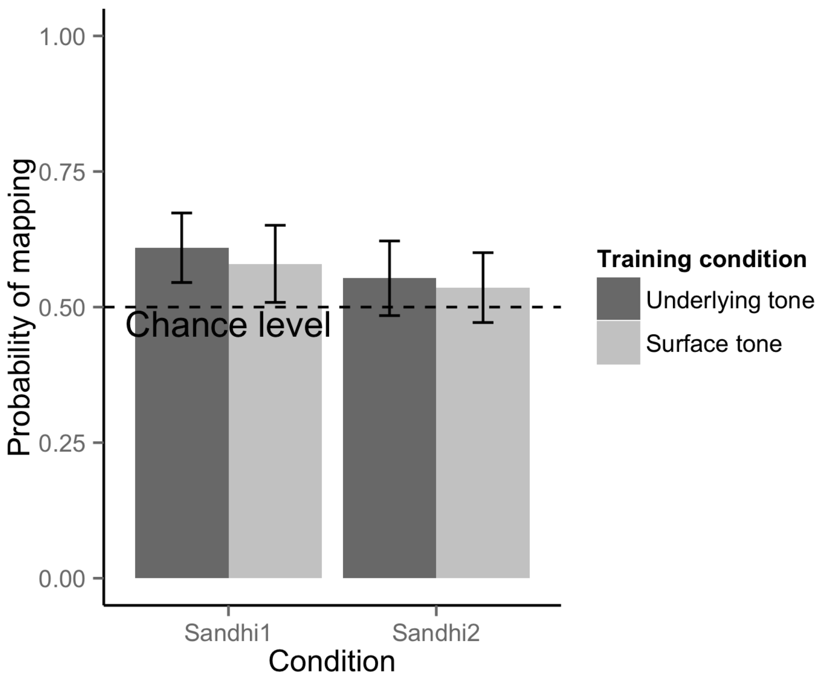

Figure 4: *Probability of mapping across conditions in the NL group. Error bars stand for ±1 standard error of mean).*

The NL group did not perform better in their surface-to-underlying mappings for Sandhi 1 the assimilatory sandhi rule. This can be interpreted as an effect of phonological uninterpretability for tones in non-tonal-language listeners, outweighing any other possible effect of factors that were manipulated in this task including the contrast between assimilatory and dissimilatory processes.

3.3.3 NJ group

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

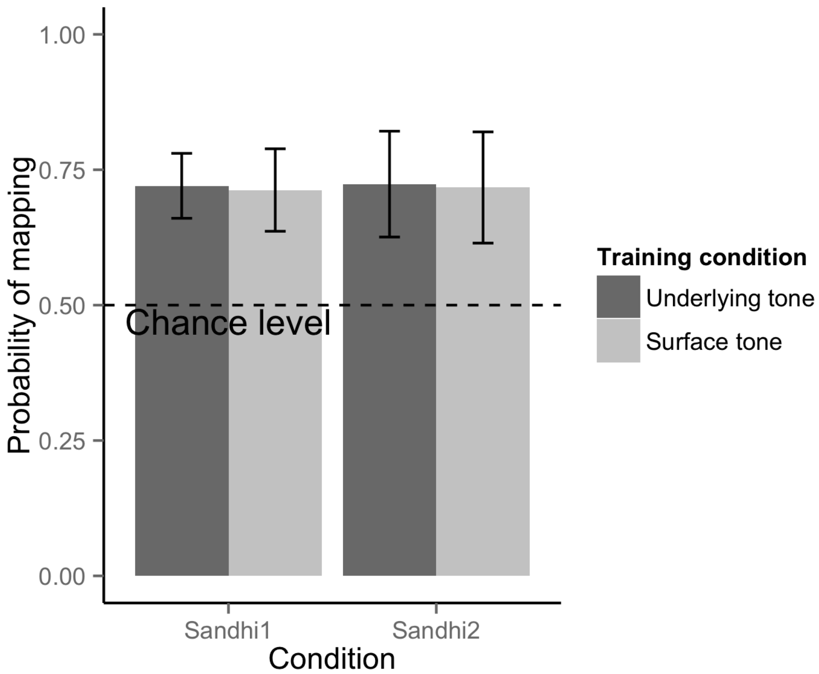

Figure 5: *Probability of mapping across conditions in the NJ group. Error bars stand for ±1 standard error of mean).*

The NJ group was the only group in this study assumed to be able to perform two types of mapping: a) mapping between the sandhied tone and the underlying tone (the underlying targeting mapping); b) mapping between the sandhied tone and the surface tone (the surface targeting mapping). Hence, two competing types of mapping were at play in this group. When the listeners were trained with the underlying tone as the target tone category, an underlying-targeting mapping allowed the NJ listeners to successfully listen through the sandhied tone to access the underlying pattern by consulting their native phonological knowledge for

Nanjing Mandarin tone sandhi rules and/or the underlying lexical representations. This underlying-targeting mapping also applied in the analysis of tones in the surface-tone condition in which the listeners learned the surface tone as the target tone category. In this condition, the NJ listeners still listened to sandhied tones at the underlying level and over-phonologized them. On the other hand, when they learned the surface tone as the target tone category, a surface-targeting mapping allowed the NJ listeners to successfully match the phonetic realization of sandhied tones with surface tones in the surface-tone condition. This is compatible with abundant existing evidence that native listeners are able to categorize speech sounds at surface level in their language extremely successfully, e.g., a segmental study by Ingram and Park (1997) observed that native Australian listeners categorized five English vowels /iː ɪ e æ aː/ with nearly 100% accuracy, and a tonal study by Francis et al. (2008) found that native Cantonese listeners categorized each one of the six Cantonese tones at above 94% accuracy with an average of 97% correct. This surface-targeting mapping also took effect when the listeners were trained with the underlying tone as the target tone category. In this condition, the listeners incorrectly mapped the sandhied tone to its surface value at the acoustic level and hence failed to perform the mapping between the sandhied tone and the underlying tone.

The results suggest that most of the NJ participants did not make a consistent choice from one of the two available types of mapping as we had predicted. Instead, they seemed to be consistently mixing the two types of mapping in this perceptual task; whenever they subconsciously adopted one mapping (e.g., in the underlying training condition: underlying-targeting mapping), the other mapping was always distracting them. Under continuous pressure from the two competing and interfering types of mapping, 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. Very few listeners consistently used a single type of mapping in doing the task, while for most NJ listeners the two types of mapping were at work competitively at the same time, although a preference for one type of mapping over the other was sometimes found (Figure 6). Previous studies investigating the underlying-targeting mapping also provide some evidence suggesting that this type of mapping is interrupted by the surface-targeting mapping by showing that the natives' surface-to-underlying mapping level is often significantly below 100%, and their response are always slower than in the control condition in which the unchanged sounds were presented. For instance, Koster (1987) found in native Dutch listeners that detection of /n/ was faster in unchanged *eten brood* 'eat bread' than in assimilated *ete*[m] *brood* cases. Moreover, the priming effect that Coenen et al. (2001) found in native listeners for German place and voice assimilations was "numerically and sometimes statistically" less for the assimilated words in viable contexts than for unchanged words, shown by longer reaction time in lexical decision and smaller effect size. Finally, Darcy et al. (2009) found that French listeners detected the target underlying word of a native voicing assimilation rule in viable context at a rate of 65% (chance level: 50%), remarkably below 96% detection rate in the no-change control condition, etc.

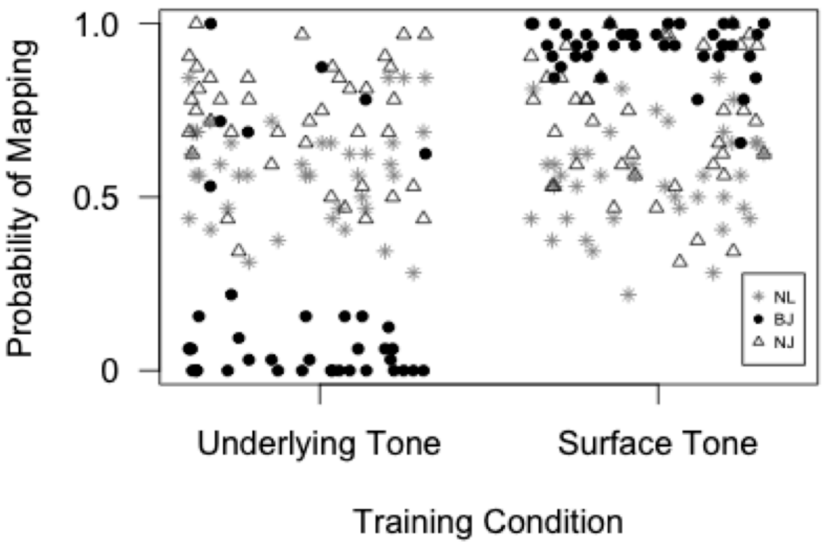

Figure 6: *Probability of mapping for the three language groups across training conditions with Sandhi 1 and Sandhi 2 combined. Each dot represents a participant.*

Although quite possibly some of the native NJ listeners may be producers of a gradient Sandhi 2 process (recall that Sandhi 2 proved to be a gradient process in X. Li and Kager (2018a) (Chapter 2)), they were nonetheless able to successfully map the categorically sandhied tone to its underlying tone (in this study Sandhi 2 was only used in its categorical form). This is probably because all listeners had been extensively exposed to both gradient and categorical forms of this sandhi process in daily life, and hence must have developed a successful mapping for all forms of this sandhi processes.

Consistent with our prediction, no significant difference between the sandhi conditions was observed in the NJ group. This is probably because the native phonology applies in both the assimilatory and the dissimilatory tone sandhi processes and always overrides any possible effect of assimilatory/dissimilatory processes in this group.

3.4 General discussion

The current study investigated whether listeners can perform mappings between surface and underlying sounds, while focusing on two aspects: a) whether the surface-to-underlying mapping depends on language-specific knowledge or is also assisted by language-general knowledge; and b) whether the mapping is influenced by assimilatory/dissimilatory types of the underlying-to-surface change, especially for naïve non-native listeners. With the tone and tone sandhi phenomena, the specific research questions were: 1) Will native tonal language listeners be the only group to successfully perform surface-to-underlying tone mapping in their native tone sandhi processes, or will a non-tonal language group, the language group most naïve of tone grammar, also be able to perform the mapping based on language-general articulatory knowledge? 2) Will assimilatory tone sandhi processes (compared to dissimilatory processes) lead to facilitated surface-to-underlying mapping for non-tonal language listeners?

With respect to research question 1), the native NJ group was the only group that successfully performed surface-to-underlying tone mapping in their native Nanjing Mandarin sandhi processes. Based on native phonological and/or lexical knowledge, they were able to undo the neutralization of contextual tonal change and perform the mapping from the sandhied tone to its corresponding underlying tone, as we predicted, in both the assimilatory and the dissimilatory processes. In addition to this predicted surface-to-underlying mapping, they were also observed to map the sandhied tone to its surface tone based on the surface value of the sandhied tone. They seemed to have consistently mixed the two ways of mapping, i.e., a) mapping the sandhied tone to the underlying tone, and b) mapping the sandhied tone to the surface tone, when perceiving native sandhied tones.

Regarding research question 2), the naïve NL listeners failed to make use of the assimilatory context to construe an articulatory-based link between the surface tone and the underlying tone in this experiment, which contradicts our prediction. Instead, they performed around chance level in both surface-to-underlying mapping and surface-to-surface mapping. This group probably managed to form fragile and unstable categories/quasi-categories for the Nanjing Mandarin target tones through a short but intensive exemplar-based tone training task, but experienced extreme difficulty in maintaining these tone categories/quasi-categories throughout the task and using them to perform the mapping task in the test session.

Adding the BJ listeners as a control group proved to be successful methodologically. On the one hand, they lack experience with the Nanjing phonological and lexical knowledge that is needed to undo the underlying-to-surface neutralization induced by the tonal context, and as such they failed to perform the mapping between Nanjing surface and underlying tones. The comparison between the performances of this group and the NJ group make it clear that the NJ listeners established the mapping based on their native language-specific phonological and/or lexical knowledge. On the other hand, the BJ listeners interpreted Nanjing Mandarin tones in terms of Beijing Mandarin tones (as argued in Section 3.4.1) at the surface level highly successfully by perceptually assimilating the Nanjing tones to their native Beijing tonal categories. This is in contrast with the NL group's failure to map the Nanjing sandhied tones to Nanjing surface tones, and adds to the evidence that the NL group failed to perform the mapping at least partly because they failed to maintain tonal categories during the experiment and use these categories in the mapping task.

The predicted-but-not-observed better mapping in the assimilatory condition by the NL group does not imply that language-general articulatory knowledge has no role to play in allowing naïve non-native listeners to perform the mapping. We suggest that the task employed in the current study may have been too cognitively challenging for the NL listeners, as the listeners were trained to learn a tone category in a short period of time (around 20 minutes) and then asked to utilize this temporally learned tone category in the mapping task. Previous studies that successfully trained non-tonal language listeners to learn tone categories mostly adopted much longer training sessions (around two or more weeks) to ensure that the categories were acquired in a reliable way (e.g., Chandrasekaran, Sampath, & Wong, 2010; Francis et al., 2008). Future studies will explore the role of assimilatory vs. dissimilatory tonal alterations in NL listeners after exposing them to a longer training time, or in a different paradigm, aiming to make the task easier for them to accomplish.

A process of contextual sound change can be categorical or gradient. When taking this categoricalness/gradience of processes into account, a number of studies can be found suggesting that naïve non-native listeners relate a surface sound and a target (underlying) sound in assimilatory processes that involve gradient changes. For example, an ERP study (Mitterer, Csépe, Honbolygo, et al., 2006) tested naïve Dutch listeners on a Hungarian liquid assimilation process (/l/ to /r/) and found that the listeners accepted the assimilated segment as its underlying counterpart more often in a viable context than in an unviable context, namely only when they heard a partially assimilated /r/. Findings like this lead us to doubt whether the facilitated surface-to-underlying mapping in the assimilatory process was not observed in the current study as a result of the categorical process, which left no phonetic residue for the naïve listeners to trace the underlying element. Future work will include gradient vs. categorical tone sandhi processes to examine if the residual phonetic traces for the underlying sound is indeed the key in allowing naïve NL listeners to successfully perform the surface-to-underlying mapping.

Chapter 4

Language generality in Dutch listeners' surface-to-underlying mapping of tones[15]

4.1 Introduction

In connected speech, the production of a speech sound is affected by the articulatory features of its neighboring sounds (e.g., Beddor & Krakow, 1999; Fowler, 1980; Hammarberg, 1976; Lahiri & Reetz, 2002; Xu, 1994), typically resulting in a *surface* (phonetic) realization of the sound that differs from its *underlying* form. This process can either be a) *assimilatory*, when a neighboring sound extends its features to the coarticulated sound, e.g., in Korean, a labial may change to a velar when immediately preceding a velar (Kim-Renaud, 1974); in English, an underlying coronal may surface as a labial when immediately followed by a labial (Gaskell & Marslen-Wilson, 1996); or b) *dissimilatory*, when the altered sound becomes less similar to a neighboring sound, e.g., in Southern Bavarian German, the liquid /r/ is converted to a non-liquid /d/ before or after an adjacent liquid /l/ (Hall, 2009). For native listeners, these surface patterns contain ambiguous information regarding the underlying form. Taking the Korean labial-to-velar assimilation rule (e.g., /ip+ko/

---

[15] Preliminary results of this chapter were presented at *the 9th International Conference on Speech Prosody*, and published in the proceedings as X. Li and Kager (2018b).

'wear and…' becomes [ikko] (Jun, 2004)) as a hypothetical case, the surface form, a velar-velar sequence, can underlyingly be either (a) a labial-velar sequence, in which the underlying labial is neutralized to a surface velar under the influence of the contextual (following) velar, or (b) a true velar-velar sequence. Native listeners have been amply demonstrated to be able to 'undo' the neutralization and *map* the surface element onto the non-surface-matching underlying element on the basis of language-specific phonological knowledge about the underlying-to-surface process and native lexical underlying representations (A. Chen & Kager, 2011; A. Chen et al., 2015; Coenen et al., 2001; Darcy et al., 2009; Gaskell & Marslen-Wilson, 1996, 1998; Gaskell & Snoeren, 2008; X. Li et al., 2016; Mitterer et al., 2013).

Regarding naïve non-native listeners, some studies have shown that they can to some extent perform surface-to-underlying mappings in assimilatory processes (e.g., Gow & Im, 2004; Mitterer, Csépe, Honbolygo, et al., 2006). However, to our knowledge, no studies exist on such mapping in dissimilatory processes. By definition, naïve non-native listeners lack knowledge of the phonological process(es) and the lexical underlying representations of a target foreign language.[16] The question arises as to how they manage to accomplish the mapping in these assimilatory processes, and what constitutes the representation onto which they map the surface form.

The difference in such mapping difficulties between assimilatory and dissimilatory processes could have partly come from different motivations underlying these processes. Assimilatory processes are generally believed to be motivated by articulatory ease (e.g., Grammont, 1933; Johnson, 1973). With regard to dissimilatory processes, the most widely acknowledged view concerning its motivation is Ohala (1993)'s proposal that dissimilatory sound changes stem from listeners' mis-attribution of surface features to coarticulation in order to maintain underlying sounds ("hyper-correction"). An alternative classic view on the motivation of dissimilatory processes is the "motor planning" view (Frisch, 2004;

---

[16] Naïve non-native listeners, theoretically, may transfer their L1 phonological knowledge in a surface-to-underlying mapping task in a foreign language. In the current study, we controlled for such transfer effects, by investigating the language-general effects with Dutch listeners, who speak a non-tonal language which is maximally remote from the target tonal language (as will be discussed at the end of this section).

Frisch et al., 2004; Garrett, 2015; Garrett & Johnson, 2012; Grammont, 1895, 1933; Tilsen, 2007). The concept of "motor planning," as Garrett & Johnson (2012, p. 59) put it, is defined as the organization of motor plans to be articulated soon by the speaker. According to this account, the motivation for dissimilatory processes lies in speaker' difficulties in planning for cooccuring sounds that are identical or similar.

Baed on the distinct motivations of assimilatory and dissimilatory processes, the language-general/specific nature of surface-to-underlying mapping in these processes can be hypothesized. The perception of acoustic events is composed of articulatory gestures, according to the Motor Theory (Liberman et al., 1967; Liberman & Mattingly, 1985) and the Direct Realist Theory (Fowler, 1986, 1996, 2006; Fowler & Brown, 2000). Based on this idea, we hypothesize that what naïve non-native listeners might be plausibly doing in this surface-to-underlying mapping process is to map a surface form onto a hypothetical underlying sound, presumably by means of implicit knowledge of articulatory gestures. Such knowledge may be generalized from experience with their native language, or alternatively may even exist independently of native language experience, e.g., by relating non-native sounds to articulatory settings by means of some form of introspection, without having experience with the articulation. This would predict that surface-to-underlying mapping in assimilatory processes may be accessible for naïve non-native listeners by means of articulatory ease, i.e., it is possibly language non-specific (the language-general mapping hypothesis for assimilatory processes). In contrast, for dissimilatory processes, the "hyper-correction" view implies that surface-to-underlying sound mapping should be only accessible for natives and inaccessible for naïve non-native listeners, because the lexical underlying representations that this view crucially refers to are unavailable for naïve non-native listeners by definition, i.e., the mapping is probably language-specific (the language-specific mapping hypothesis for dissimilatory processes). Nevertheless, predictions derived from the "motor planning" view are unclear and inconsistent regarding the language specificity/generality of the mapping for dissimilatory processes. For this reason, we will only test the language-specific mapping hypothesis clearly predicted by the "hyper-correction" account for dissimilatory processes.

As a specification of the language-general mapping hypothesis for assimilatory processes, a surface-to-underlying mapping for an assimilatory process may be available to naïve non-native listeners only if the underlying representation is to some extent acoustically recoverable from the surface (the gradient-based language-general mapping hypothesis for assimilatory processes). A process of contextual sound change could occur at different degrees. The process can be *categorical*, in which case the articulatory gesture or characteristic of the underlying element is *completely lost* in the production of the surface form and the derived sound has become *neutralized* with another sound. For example, in the Korean labial-to-velar assimilation, the surface labial resembles un underlying velar with extensive gestural overlap (Jun, 2004). Alternatively, the process can also be *gradient*, in which case articulatory gesture of the underlying sound is only *partially lost* from the surface form. For example, in English coronal place assimilation, assimilated coronals were found to have traces of the coronal segments in formant features (Gow, 2003). Hence, we may assume that *non-neutralizing* assimilatory processes, which involve gradient underlying-to-surface sound changes and leave some residual phonetic traces of the underlying sound in the surface form, would be more likely to allow naïve non-native listeners to relate the surface and the underlying sounds.

Empirical evidence can be found to support the language-general mapping hypothesis for assimilatory processes, however the various conclusions diverge as to whether this mapping crucially relies on gradience, i.e., the recoverability of acoustic traces of the underlying sound remaining at the surface.

Most studies found that naïve non-native listeners were able to perform the mapping between the surface and the underlying sounds successfully only for gradient assimilatory processes in a context-sensitive way. For instance, Gow and Im (2004) observed that naïve English listeners performing a mapping task for a gradient Hungarian voicing assimilation process (e.g., *oros dInAstiA* 'Russian dynasty' becomes *oro*[z] *dInAstiA*) were facilitated in monitoring a target (underlying) sound in its surface form in a 'viable context,' which allowed for the assimilatory change, as compared to an 'unviable context,' which did not permit the assimilatory change; whereas naïve English listeners did not show such facilitating effects in the Korean categorical labial-to-velar assimilation process (e.g., /gom+gæʃi/ 'bear-like' becomes [goŋgæʃi]). Mitterer et al. (2013) similarly observed for the Korean categorical labial-to-velar assimilation process that naïve Dutch and English groups of listeners both failed to categorize the derived velar as a target (underlying) labial more in the viable context, compared to natives. An ERP study (Mitterer, Csépe, Honbolygo, et al., 2006) tested naïve Dutch listeners on a Hungarian liquid assimilation process (/l/ to /r/) (e.g., /knɑl+ro:t/ 'vivid red' becomes [knɑrro:t], an application of Hungarian liquid assimilation to Dutch words) and found that the listeners accepted the assimilated segment more frequently as its underlying counterpart in a viable context than in an unviable context when they heard a gradiently assimilated /r/, but not when they heard a categorically assimilated /r/. A single study suggests that naïve non-native listeners relate the surface and the target (underlying) sounds in categorical assimilation processes as well. Darcy et al. (2009) tested naïve French listeners on a categorical English place assimilation process (e.g., /wet+pænts/ becomes [weppænts]) as well as naïve English listeners on a categorical French voicing assimilation process (e.g., /gid+savɑ̃/ 'learned guide' becomes [gitsavɑ̃]) using a word detection experiment (details of this paradigm are presented in Section 4.2), and observed that both groups of listeners detected the target underlying word more successfully in viable than in unviable contexts, despite the use of neutralized phonemic contrasts in both the processes.

Regarding dissimilatory processes, no previous study has investigated whether naïve non-native listeners could perform a mapping between surface sound and underlying sound, as far as we are aware. Whether the surface-to-underlying mapping in dissimilatory processes is inaccessible to naïve non-native listeners, as the "hyper-correction" theory (Ohala, 1993) implies, remains unknown.

The current study aims to investigate a) whether assimilatory and dissimilatory processes differ in their recoverability of surface-to-underlying mapping for naïve non-native listeners. It also investigates b) whether gradient and categorical sound changes further lead to any difference in surface-to-underlying mapping for naïve non-native listeners. For research question a), we predict that for naïve non-native listeners, surface-to-underlying mapping is only possible for assimilatory processes (the language-general mapping hypothesis for assimilatory processes) which are articulatorily motivated in terms of language-unspecific knowledge. For research question b), our hypothesis is that surface-to-underlying mapping may be facilitated only in those assimilatory processes that involve gradient changes (the gradient-based language-general mapping hypothesis for assimilatory processes).

In order to subject the language-general surface-to-underlying mapping hypothesis for assimilatory processes to the strongest possible test, we intend to keep the non-native listener group as naïve as possible. As discussed earlier, we assume that for naïve non-native listeners, knowledge of (universal) articulatory gestures enables them to bridge the surface sound and the hypothetical underlying sound. For example, American English listeners, despite having no experience with Tamil sounds, parsed the place of articulation on a following stop consonant to its precursor consonant (Viswanathan, Magnuson, & Fowler, 2010). Studies like this lead us to assume that this general articulatory knowledge can be extended to lexical tone and tone sandhi phenomena. By using the phenomena of lexical tone and tone sandhi to investigate the mapping issue, we will be able to test a group of naïve non-native listeners that is maximally remote from the native language group, namely a non-tonal language group without any previous exposure to tones, hence guaranteed to have no articulatory experience with tone sandhi. Tone sandhi processes can be assimilatory and dissimilatory. For example, Zhenjiang Mandarin (Qiu, 2012) has an assimilatory tone sandhi rule turning a high falling T1 immediately before another high falling T1 into a high level T4, with tone offset raised to the same pitch height of its context tone onset, written as HL.HL → HH.HL (e.g., *jin*T1 *shan*T1 金山 'gold mountain' → [jinT4 shanT1进山] 'to get into a mountain'). The T3 sandhi rule in Beijing Mandarin is an example of the dissimilatory tone sandhi process (Cheng, 1973; Yip, 1980, 2002): a low T3 (the distinctive feature of lexical Mandarin Tone 3 is "low") immediately preceding another low T3 changes to a low-rising T2, transcribed as LL.LL → LH.LL (e.g., *mai*T3 *ma*T3 买马 'to buy a horse' → [maiT2 maT3 埋马] 'to bury a horse') (e.g., Yip, 1980), with the altered tone offset deviating from the subsequent context tone onset. Tone sandhi processes also fall apart into categorical and gradient types. For example, Taiwan Southern Min has a tone sandhi process which changes a high level *yinping* (in traditional tone classification of this dialect) categorically to a mid level *yangqu* before a low rising *yangping* (Myers & Tsay, 2008). Tonal coarticulation processes involving gradient underlying-to-surface changes are also prevalent, e.g., in Tianjin Mandarin, a high-falling T4 preceding a mid-falling T1 turns into a high-rising tone which resembles a lexical mid-rising T2 but keeps its original high onset (Q. Li & Chen, 2016).

In the current study, an artificial tonal language was constructed as the target language, implementing two dimensions of underlying-to-surface tonal changes: a) assimilatory and dissimilatory underlying-to-surface changes and b) gradient and categorical underlying-to-surface changes. Dutch listeners were used as the naïve non-tonal language listener group in this study.

4.2 Experiments

The Word Detection task (Darcy et al., 2009) was adopted in the current study. In Darcy et al. (2009)'s study, a target word was presented and followed by a sentence containing the target, surfacing with place assimilatory changes and voicing assimilatory changes. Native and naïve non-native listeners were requested to judge whether the target presented was the same as in the sentence. They successfully used this paradigm to test the listeners on their surface-to-target (underlying) mapping in the two assimilatory rules. Guided by their study, we applied this paradigm in the case of a tone sandhi process to test whether naïve Dutch listeners can relate a surface sandhied tone with a target (underlying) tone. To simplify the task for Dutch listeners, we embedded the target only in its tonal context instead of in a sentence as Darcy et al. (2009) did. In the current study, target words (in their underlying tonal shape) were presented auditorily and followed by test items containing the target word (now appearing in a gradient versus categorical sandhied tone shape, and embedded either in an assimilatory or dissimilatory context). Participants had to decide whether they detected the target word in the test items. A comparison of detection performance of target words in assimilatory versus dissimilatory contexts and of test words containing gradient versus categorical sandhied tones will demonstrate how Dutch listeners are able to make use of the two factors manipulated in a surface-to-target (underlying) tone mapping task.

The underlying tone is implemented in the current study as a phonetic tone presented to the listeners. One might argue that for naïve non-native listeners, the assumed mapping between the sandhied tone and the given underlying tone is purely mapping at surface level because the listeners lack underlying forms and will hence only take a given underlying tone as its surface value. This view, nonetheless, would not change our hypotheses of (gradient-based) language-general mapping for assimilatory processes and language-specific mapping for dissimilatory processes because the naïve non-native listeners, though possibly taking the underlying tone as its surface value, are still confronted with the task of mapping between this tone and a phonetically variant - the sandhied tone. Undoing this type of form variation should involve the same perceptual knowledge on the part of the listener as surface-to-underlying mapping. In this study, we refer to this mapping process by naïve non-native listeners as surface-to-underlying mapping for convenience.

4.2.1 Experiment 1

Experiment 1 included categorical underlying-to-surface changes, either assimilatory or dissimilatory. Based on the language-general surface-to-underlying mapping hypothesis for assimilatory processes and language-specific mapping hypothesis for dissimilatory processes, we predicted that naïve Dutch listeners would perform surface-to-target (underlying) mapping more successfully in an assimilatory process than in a dissimilatory process in the artificial tonal language.

4.2.1.1 Stimuli

A pair of tone sandhi rules involving categorical underlying-to-surface tone changes and differing in assimilatory/dissimilatory nature were created in an artificial tonal language. Sandhi 1: 44.24 → 42.24[17]; Sandhi 2: 44.42 → 42.42. In Sandhi 1, the offset of a high-level 44 in the first syllable is lowered to the same

[17] The digits indicate the rough heights of pitch using Chao (1930)'s 5 level representation, with 1 representing the lowest pitch level in the listener's pitch range and 5 the highest; the dot is used to separate tones.

pitch as the onset of the following 24, which can be viewed as a categorical tonal assimilatory process; Sandhi 2 is a categorical dissimilatory process, since the offset of a high-level tone in the first syllable deviates far from the subsequent onset of a high-falling 42 instead of approaching it. The two tone sandhi rules both involved a tone change from an underlyingly high level tone 44 to a surface high-falling tone 42. The two artificial tone sandhi rules were used as the base for creating stimuli in this experiment.

Four target words carrying the underlying tone 44 were selected: /ba44/, /bi44/, /de44/, and /go44/. They were all monosyllabic words beginning with a voiced stop and ending with a vowel. The target words were spoken naturally by a 27-year-old female Beijing Mandarin speaker. The tokens were digitized at 16 kHz and 16 bits. They were normalized in duration (450 ms) and in intensity.

Four matched test words bearing the surface tone 42 ([ba42], [bi42], [de42], [go42]) and two context words [du24] and [du42] were recorded by another 28-year-old female Mandarin speaker. The test words and the context words were normalized in duration (450 ms) and in intensity before concatenation. Each of the 4 test words was concatenated with each of the 2 context words with a 70 ms silent portion in between,[18] creating 8 disyllabic test items (e.g., [ba42 du24] (assimilatory condition)/[ba42 du42] (dissimilatory condition)). Figure 1 and 2 shows pitch contours of the tones 42.24 and 42.42, respectively. The test items were repeated 8 times using software and then randomly assigned into 2 blocks.

In order to draw participants' attention to the tonal shape of words in the experiment and to prevent them from simply agreeing when detecting the target segmental shape in a test item, 8 additional test items were constructed as practice trials. The practice trials contained a tonally-changed test item (e.g., [ba24 du22]) and a tonally unchanged test item (e.g., [ba44 du24]) for each of the 4 target words. None of the practice items carried the same tone pattern as the 2 experimental conditions.

[18] The 70 ms silent portion was inserted after all the 3 Dutch listeners in a pilot study reported unnaturalness of test words adjacently concatenated with context words. After insertion of the silent portion, the Dutch listeners reported the sounds to be natural and to contain no obvious break in between.

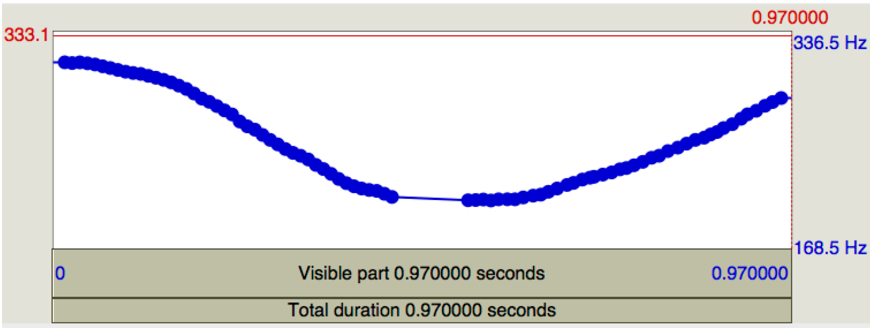

Figure 1: Pitch contour of the tone 42.24 (lefthand: tone 42; righthand: tone 24).

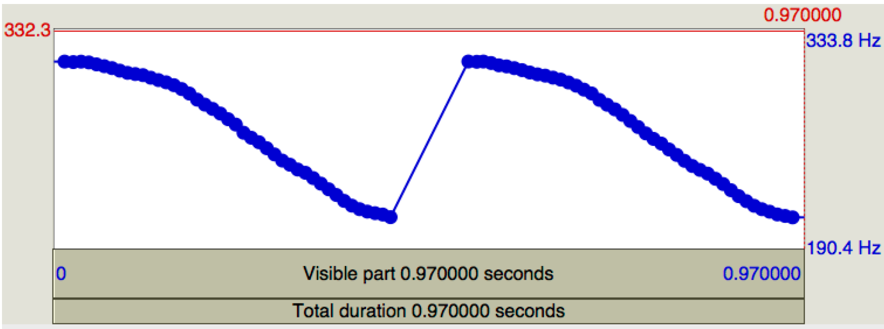

Figure 2: Pitch contour of the tone 42.42 (lefthand: tone 42; righthand: tone 42).

All the recording and editing of the words used in the experiment were done using Audacity and Praat (Boersma & Weenink, 2009).

4.2.1.2 Participants

Thirty Dutch participants (15 males and 15 females) with no previous exposure to a tonal language were recruited in this experiment. All of them were aged between 18 and 30, and all had self-reported normal hearing. Every participant completed the two conditions (assimilatory/dissimilatory) in one experiment. The order of trials within a block and the order of the blocks were randomized for each participant.

4.2.1.3 Procedure

The participants were tested in a sound-proof booth in the Phonetic Lab of the Utrecht Institute of Linguistics. Experimental trials were pairs beginning with the presentation of the monosyllabic target word (e.g., [ba44]), followed after 1000 ms of silence by a disyllabic test item (e.g., [ba42 du24]/ [ba42 du42]). The inter-trial interval was 4000 ms. Participants were instructed to press a "yes" button when they thought the target word appeared in the test item with the same melodic shape, and to press a "no" if otherwise[19] (for the full instructions, see Appendix G). In each trial they were allowed a total of 2000 ms after the offset of the disyllabic word to make their response. Missed trials were repeated at the end of the trials so that the participants' responses to all the trials were collected. The experiment conditions are shown in Table 1.

Table 1. *Conditions in Experiment 1.*

	Target (presented first)	Test item (presented afterwards)	
Tones	44	42.24	Assimilatory condition
		42.42	Dissimilatory condition

In the practice session preceding the test session, practice trials were played to the participants in a loop. Visual "correct/incorrect" feedback was provided after participants made each response. The practice session automatically ended when a listener performed 7 trials in a row with 2 or fewer errors, thus judged to have switched their attention to tonal shape instead of segmental shape of the words.[20] They then proceeded to the test session. During the test session, responses were collected without giving feedback.

[19] For the right-handed listeners, the right button on the buttonbox was set as the "yes" button; for the left-handed listeners, the left button was the "yes" button.

[20] The criterion for having switched attention to the tonal shape was settled according to the results of a pilot experiment with 5 NL listeners. The NL listeners made many mistakes in the beginning and gradually made fewer mistakes until finally passing the criterion, generally within 10 minutes.

The experiment was conducted via the program Zep Experiment Control Application (Veenker, 2017). Both the detection (yes/no) and reaction time for each test item were recorded. A higher detection rate would suggest a more frequent surface-to-underlying mapping; shorter reaction time for "yes" responses would suggest that listeners are more certain about the mapping. Both the parameters are considered to assess the mapping performance.

4.2.1.4 Results and discussion

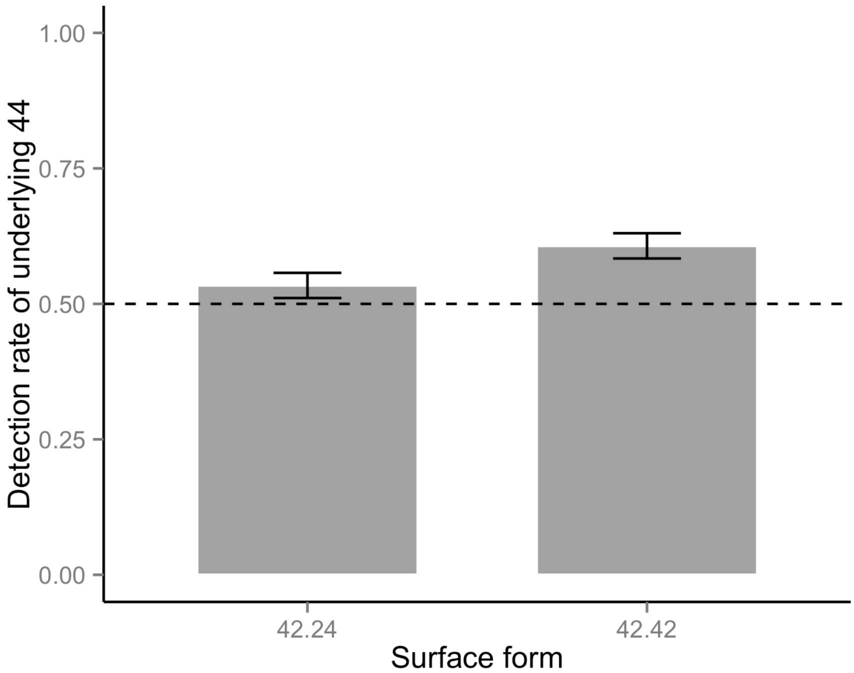

Figure 3: *Detection rate of the underlying 44 in the surface forms 42.24 (assimilatory condition) and 42.42 (dissimilatory condition) in the Dutch participants. Error bars stand for ±1 standard error of the mean.*

Before data analysis, raw reaction times were square-root (sqrt) transformed to achieve a normal distribution. Trials with response times that were beyond 2 standard deviations from the grand mean were excluded as outliers.

A generalized linear mixed-effects model was constructed to analyze the detection data in R (R Core Team, 2014) using the *glmer* function in the *lme4* package (Bates et al., 2014). The dependent variable in the analysis was the detection value (yes/no). PARTICIPANT and ITEM were included as random effects. The fixed effect of SANDHI (or surface form) (assimilatory 42.24/dissimilatory 42.42) was added to the model. By log-likelihood model fit comparison, the fixed effect of SANDHI proved to significantly improve the fit of the model ($\chi^2(1) = 10.72$, $p = .001$ **). The model revealed significantly lower detection of the underlying tone 44 in the surface form 42.24 than in the surface form 42.42 (Estimate = 0.33, $z = 3.28$, $p = .001$ **). Figure 3 depicts the detection rate across the two conditions.

For reaction time data, only "yes" responses (which indicated that the listeners performed the mapping) were included in the data analysis. A linear mixed-effects model was constructed to analyze the reaction time data in R (R Core Team, 2014) using the *lmer* function in the *lme4* package (Bates et al., 2014). The dependent variable in the analysis was the sqrt reaction time value. PARTICIPANT and ITEM were included as random effects. The fixed effect of SANDHI (assimilatory 42.24/dissimilatory 42.42) was added to the model. By log-likelihood model fit comparison, the fixed effect of SANDHI proved to significantly improve the fit of the model ($\chi^2(1) = 8.48$, $p < .01$ **). The model revealed significantly shorter reaction time when the listeners performed the mapping in the surface form 42.42 than in the surface form 42.24 (Estimate = -1.32, $t = -2.92$, $p < .01$ **). Figure 4 depicts the reaction time across the two conditions.

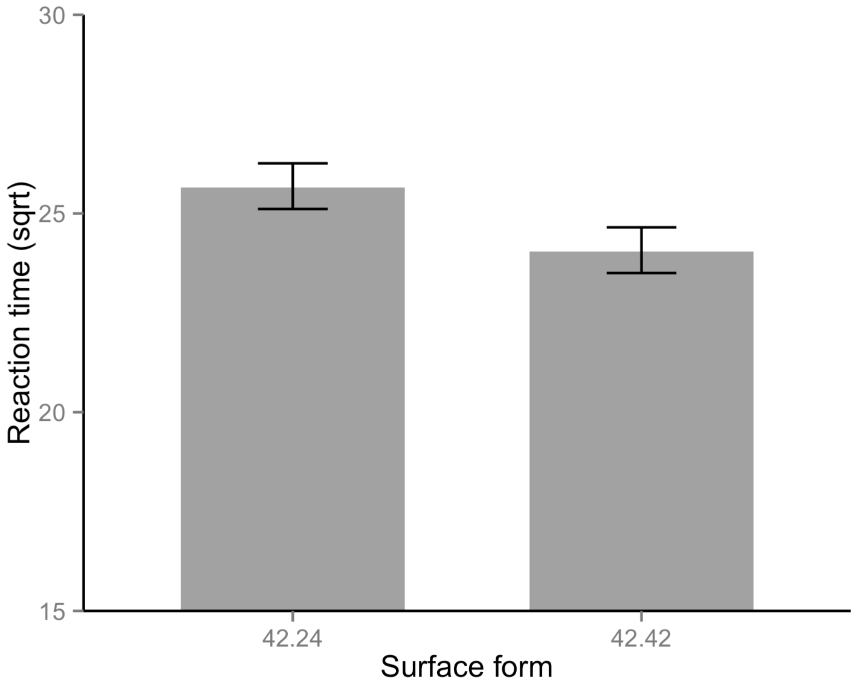

Figure 4: *Reaction time (sqrt) of the Dutch listeners' mapping in the surface forms 42.24 (assimilatory condition) and 42.42 (dissimilatory condition) ("yes" responses only). Error bars stand for ±1 standard error of the mean.*

It turned out that the Dutch listeners demonstrated a word detection pattern opposite to our hypothesis of better mapping in the assimilatory condition, derived from the articulatory-ease motivated theory for assimilatory processes (e.g., Grammont, 1933; Johnson, 1973) and the "hyper-correction" theory (Ohala, 1993) for dissimilatory processes. Instead, the listeners detected the underlying tone 44 more frequently and faster in the 42.42 surface form, which involved a categorical dissimilatory underlying-to-surface tonal change from the underlying form (44.42 → 42.42), compared to the 42.24 surface form, which included a discrete assimilatory underlying-to-surface tonal change (44.24 → 42.24). The cause of the unexpected better mapping in the dissimilatory condition observed in Experiment 1 will be discussed together with the results of Experiment 2 in the next section.

The finding that the surface-to-target (underlying) mapping did not benefit from the assimilatory condition in Experiment 1 might be due to the categorical change from the underlying tone 44 to the surface tone 42 which, according to the gradient-based mapping hypothesis for assimilatory processes, left no acoustic traces of the target (underlying) tone in the surface form for the Dutch listeners to perform the mapping. Experiment 2 aims to investigate the same mapping when the change becomes more gradient.

4.2.2 Experiment 2

Experiment 2 was designed to examine whether a gradient (rather than categorical) change from the underlying tone to the surface tone would elicit more successful surface-to-target (underlying) tone mapping in the assimilatory condition. This was implemented by changing the surface tone to 43, i.e., a less steeply falling tone that is intermediate between 42 (the surface tone in Experiment 1) and 44 (the target (underlying) tone). We hypothesized that more successful surface-to-underlying mapping might be observed in the assimilatory condition when the change is more gradient.

4.2.2.1 Stimuli

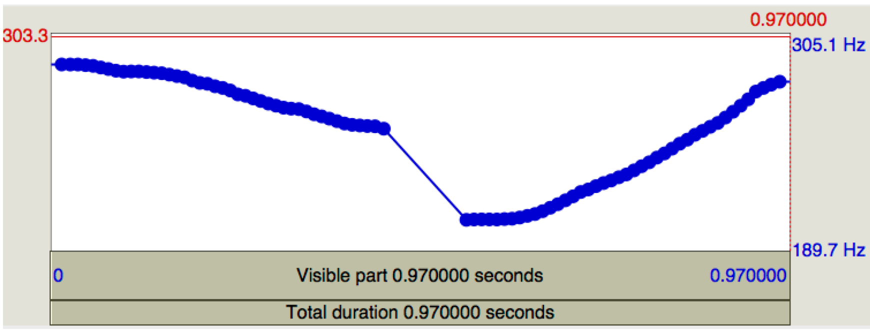

Figure 5: Pitch contour of the tone 43.24 (lefthand: tone 43; righthand: tone 24).

The same stimuli as in Experiment 1 were used in the current experiment, except that the tones in the monosyllabic test words were changed to 43, and accordingly the tone sequences on the disyllabic test items were altered to 43.24 and 43.42. The target tone presented prior to the disyllabic word was 44 once again. The tone contour 43 was constructed by taking the perceptual midpoint between 44 and 42. The perceptual midpoint was found to be step 3 out of 7 steps interpolated between 44 and 42 linearly along the Hertz scale, based on a discrimination experiment that we conducted[21]. It is referred to as 43 for convenience. Figure 5 and 6 demonstrate pitch contours of the tones 43.24 and 43.42.

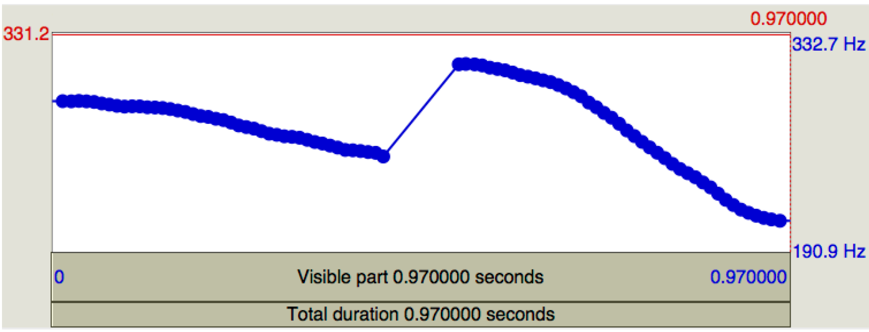

Figure 6: Pitch contour of the tone 43.42 (lefthand: tone 43; righthand: tone 42).

4.2.2.2 Participants

Another thirty Dutch participants (15 males and 15 females) were recruited for Experiment 2. They met the same criteria as in Experiment 1.

---

[21] In the discrimination task, three equidistant steps were interpolated between tone 44 and tone 42 for syllables /ba/, /du/ and /si/. Each syllable was made as a block. In each block, 25 (5^2) comparisons (trials) were made between the five stimuli including the endpoints 44 and 42, with counterbalanced order of stimuli presentation. The intra-trial interval was 250ms; the inter-trial interval was 3000ms. In each trial, participants heard a pair of stimuli and determined whether the stimuli sounded the same or different. Eight participants took part in the experiment. Based on the results, the perception midpoint was found to be between the step 1 and 2 out of the 3 steps interpolated, which would be around step 3 if 7 steps were interpolated between the endpoints.

4.2.2.3 Procedure

The procedures were the same as in Experiment 1. The experimental conditions are shown in Table 2.

Table 2. *Conditions in Experiment 2*

	Target (presented first)	Test item (presented afterwards)	
Tones	44	43.24	Assimilatory condition
		43.42	Dissimilatory condition

4.2.2.4 Results and discussion

Reaction times were first sqrt-converted, and trials with response times that were beyond 2 standard deviations from the grand mean were excluded as outliers.

A generalized linear mixed-effects model was built following the same steps in Experiment 1 to analyze the detection data. The fixed effect of SANDHI did not significantly improve the fit of the model ($\chi^2(1) = 0.4159$, $p > .5$). The model revealed no significant difference between the detection rates of the 43.24 and the 43.42 conditions (Estimate = -0.06, z = -0.65, $p > .5$), as shown in Figure 7.

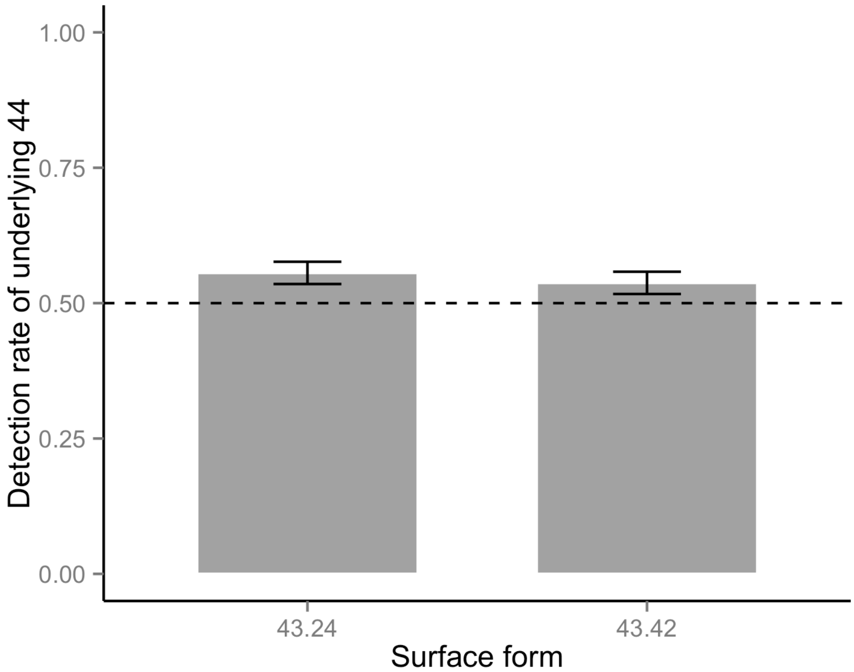

Figure 7: *Detection rate of the underlying 44 in the surface forms 43.24 (assimilatory condition) and 43.42 (dissimilatory condition) in the Dutch participants. Error bars stand for ±1 standard error of mean.*

Only "yes" responses (which indicated that the listeners performed the mapping) were included in the data analysis for the reaction time data. A linear mixed-effects model was constructed following the same steps in Experiment 1. The fixed effect of SANDHI significantly improved the fit of the model ($\chi^2(1)$ = 12.31, p < .001 ***). It showed significantly faster reaction time when the listeners performed the mapping in the surface form 43.24 than in the surface form 43.42 (Estimate = 1.60, t = 3.53, p < .001 ***), as shown in Figure 8.

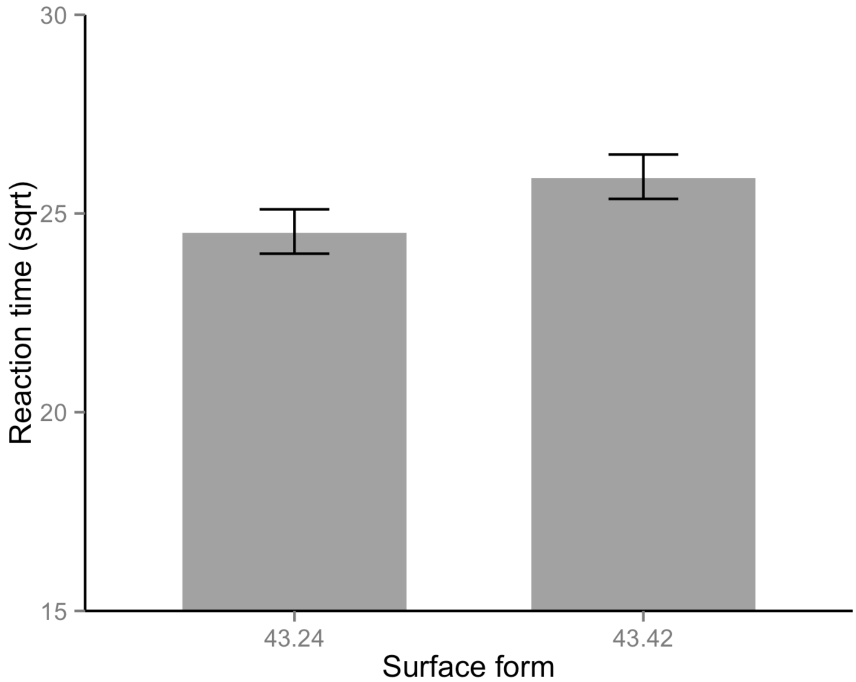

Figure 8: *Reaction time (sqrt) of the Dutch listeners' mapping in the surface forms 43.24 (assimilatory condition) and 43.42 (dissimilatory condition) ("yes" responses only). Error bars stand for ±1 standard error of the mean.*

The results show that the Dutch listeners did not perform surface-to-target (underlying) tone mapping more frequently in the assimilatory condition when the tonal change from the underlying tone to the surface tone was more gradient, though a noticeable trend of higher mapping rate in the assimilatory condition compared to the dissimilatory condition was observed. However, reaction time results showed that the Dutch listeners responded faster in surface-to-target (underlying) tone mapping in the assimilatory condition, suggesting that the listeners were more confident about their responses when performing the mapping in this condition. In a phoneme monitoring study targeting context effects by naïve listeners (Gow & Im, 2004), facilitated monitoring of the underlying sound in the viable context was also only observed with reaction time data, whereas the effect did not occur in the accuracy data that they measured. The mapping rate in

Gow and Im (2004) was higher (around 85%) than in the current study (around 55%). This difference with the current study may be due to their use of Hungarian sounds to test American English listeners, using stimuli arguably more familiar to the listeners than tones are to the Dutch listeners. The use of tone, and tone sandhi phenomena in particular, may have caused the Dutch listeners to be less sure about their responses in general, and hence they showed an overall decrease in the accuracy of their mapping in both the assimilatory and dissimilatory conditions in this study. Recall from Section 4.1 that in order to subject the (gradient-based) language-general mapping hypothesis for assimilatory processes to the strongest test, it was decided to expose naïve non-native listeners to speech phenomena that are maximally distant from their native language, i.e., tones. In retrospect, auch an emphasis on maximally distant non-native speech sounds may have had an intrinsic downside, as Dutch listeners' problems with processing tones probably caused an overall decrease in their mapping rate.

A comparison between the results of Experiment 1 and Experiment 2 provides some insight as to why the Dutch listeners, opposite from what we had hypothesized, showed a mapping advantage for the dissimilatory condition 42.42 over the assimilatory condition 42.24 in Experiment 1. The categorical assimilatory 42.24 condition and the categorical dissimilatory 42.42 condition differ in the f0 agreement/disagreement between the offset of the surface tone 42 and the onset of its following context. The f0 disagreement in the 42.42 sequence (dissimilatory) probably caused a facilitation in the segmentation between the surface tone 42 and its following context 42 and hence created a more transparent connection between the surface tone and the target (underlying) tone for the Dutch listeners. Although the current experimental setting included a 70 ms physical pause between the surface tone and its following tonal context which exceeds the subliminal pause duration (25 ms) that is sufficient to bring about the segmentation of a segmental sequence (Peña, Bonatti, Nespor, & Mehler, 2002), locating a boundary in a tone sequence in spite of the presence of a brief pause may have been difficult for the Dutch listeners, so that f0 disagreement in the dissimilatory condition may have significantly enhanced their segmentation.

The role of dissimilation as a potential segmentation marker was already noticed by Johnson (1973). He observed that cross-linguistically, dissimilation phenomena occur mostly across concatenated morphemes, and suggested that dissimilation processes apply to "preserve the distinctiveness of the stem-affix relationship" and to "make stem boundaries more prominent," which implied segmentation between the stem and the affix. In our Experiment 1, the Dutch listeners may have also more easily segmented the surface tone 42 from its context in the categorical dissimilatory 42.42 condition because of the f0 disagreement between the surface tone and its tonal context, whereas the f0 agreement between the surface tone 42 and the contextual tone 24 in 42.24 (assimilatory condition) may have led to more difficulty when segmenting the surface tone 42. Evidence from previous studies suggests that f0 disagreement can be used to suggest boundaries. For example, H.-Y. Lin and Fon (2011) found a clear effect of pitch reset on the detectability of discourse boundaries and on subjects' ranking of boundary hierarchy; Brugos and Barnes (2014) found that silent intervals bounded by tokens of widely differing pitch are heard as longer than those bounded by tokens closer in pitch.

Alternatively, a different type of enhanced segmentation cue could have come from the double occurrence (repetition) of tones. In the 42.42 (dissimilatory) condition of Experiment 1, the repetition of tone 42 may have contributed to an enhanced segmentation of the surface tone 42 from its context. Many studies have reported an under-representation of repeated sounds in the phonological forms of morphemes. For example, Monaghan and Zuidema (2015) noted an underrepresentation of syllable repetition in the lexicons of many languages. Boll-Avetisyan and Kager (2014) empirically proved that the co-occurrence of pairs of identical consonants in continuous speech is used as a segmentation cue by Dutch listeners. Restrictions on repetition of identical tones is widespread in the lexicon and surface phonology of many languages, e.g., the massive evidence for OCP-high tone in Bantu languages (Meyers, 1997). These studies lead to the question of whether sound repetition as a segmentation cue also holds for tones.

In Experiment 2, the gradient assimilatory 43.24 condition and the gradient dissimilatory 43.42 condition did not differ in the f0 agreement/disagreement between the offset of the surface tone 42 and the onset of its following context; accordingly, repetition of identical tones was avoided in the surface tone sequence. Hence, it may be assumed that in Experiment 2, the segmentation of the surface tone sequence in the two conditions was equally clear and unclear to the Dutch participants. As segmentation differences arguably played no role in Experiment 2, the hypothesized better mapping in the assimilatory (compared to dissimilatory) condition should have re-emerged. We indeed observed evidence of this better mapping, as the Dutch listeners provided faster responses when performing the mapping in the assimilatory (compared to dissimilatory) condition. Furthermore, the mapping in this condition tended to be more frequent in the detection rate results, although not enough to be statistically significant.

It therefore remains unclear based on the current results whether enhanced segmentation for the Dutch listeners in the 42.42 surface form (dissimilatory condition) in Experiment 1, compared to the 42.24 surface form in the same experiment, was due to a) f0 disagreement on the tone boundary or b) repetition of the 42 tone. Experiment 3 was thus conducted to disentangle these two possible effects.

4.2.3 Experiment 3

As a variation on Experiment 1, Experiment 3 added a third tone condition 42.44322 (details in next section) as the surface form with f0 disagreement on the tone boundary intact (+ f0 disagreement) while tone repetition was removed (- repetition) in order to allow comparisons with the 42.42 condition (+ f0 disagreement, + repetition) and the 42.24 condition (- f0 disagreement, - repetition). The third tone condition agreed with the 42.24 condition in non-repetition while differing from it in f0 agreement/disagreement; it agreed with the 42.42 condition in f0 disagreement while differed from it in repetition/non-repetition.

4.2.3.1 Stimuli

The stimuli from Experiment 1 were included in the current experiment, and a new test item condition 42.44322 was added. The new condition 42.44322 carried two distinct (non-repetitive) tones on the two syllables, while retaining the f0 disagreement between the surface tone and its contextual tone. The tone shape

44322 was created by editing the pitch tier of the 42 tone and then resynthesizing using the *resynthesis (overlap and add)* function in Praat (Boersma & Weenink, 2009) while keeping the duration intact as 42, as shown in Figure 9.

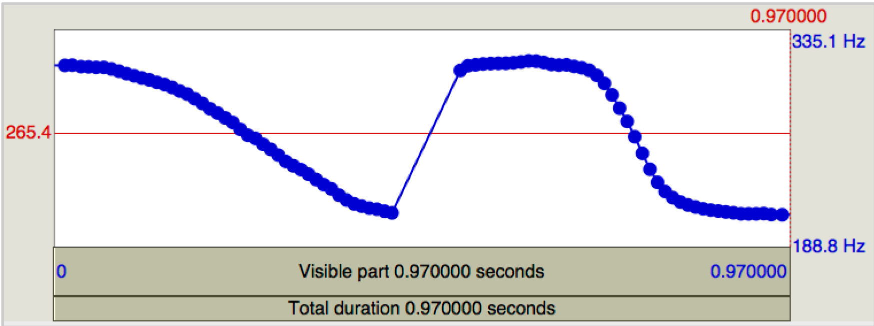

Figure 9: Pitch contour of the tone 42.44322 (lefthand: tone 42; righthand: tone 44322).

Based on our observation of better mapping performance, i.e., higher detection rate of the target and shorter reaction time in the 42.42 condition than in the 42.24 condition in Experiment 1, the predictions for the current experiment are as follows: a) if the 42.44322 surface form elicits the same detection performance as 42.24 with significantly lower detection rate and longer reaction time than 42.42, then it should be tone repetition (in 42.42) that is making the segmentation easier for the Dutch listeners; b) if 42.44322 elicits the same detection performance as 42.42 with significantly higher detection rate and shorter reaction time than 42.24, then f0 disagreement (in both 42.44322 and 42.42) should be the main cause of easier segmentation by the Dutch listeners; c) if the detection performance in 42.44322 is found intermediate between 42.24 and 42.42 and significantly different (in detection rate and reaction time) from both of them, then both f0 disagreement and repetition are contributing to easier segmentation.

4.2.3.2 Participants

Thirty Dutch participants (15 males and 15 females) were recruited for the current task. They met the same criteria as in Experiment 1 and 2, but had not participated in these previous experiments.

4.2.3.3 Procedure

The procedure was the same as in Experiment 1 and 2. The experiment conditions are shown in Table 3.

Table 3. *Conditions in Experiment 3*

	Target (presented first)	Test item (presented afterwards)		
Tones	44	42.24	- f0 agreement	- repetition
		42.42	+ f0 agreement	+ repetition
		42.44322	+ f0 agreement	- repetition

4.2.3.4 Results and discussion

Reaction times were first sqrt-converted, and trials with response times that were beyond 2 standard deviations from the grand mean were excluded as outliers.

A generalized linear mixed-effects model was built following the same steps in Experiment 1 and 2 to analyze the detection data. The model revealed that TONE (42.24/42.42/42.44322) was a significant predictor of the model ($\chi^2(2) = 7.18$, $p = .028$ *). Post-hoc paired comparisons between the conditions were conducted using the *glht* function in the *Multcomp* package (Hothorn, Bretz, & Westfall, 2008) in R. The comparisons revealed: a) significantly higher detection rate in the 42.42 condition than the 42.24 condition (Estimate = 0.26, $z = 2.57$, $p = .03$ *); b) slightly but not significantly higher detection rate in the 42.44322 condition than in the 42.24 condition (Estimate = 0.20, z = 1.95, p = .12); c) no difference between the detection rates of the 42.44322 and the 42.42 conditions. (Estimate = -0.06, z = -0.62, p = .81). Figure 10 shows the detection rate across the three conditions.

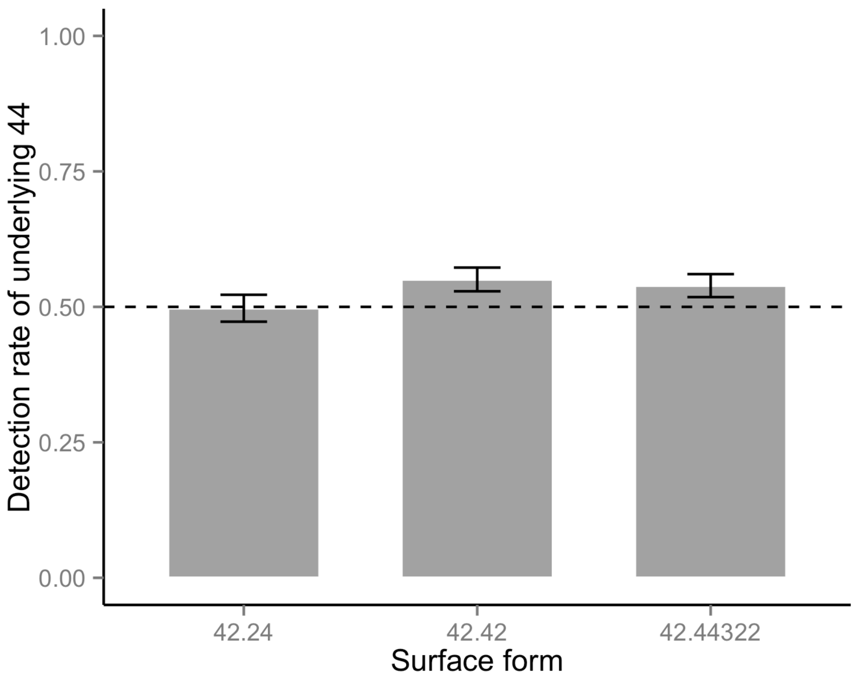

Figure 10: *Detection rate of the underlying 44 in the surface forms 42.24, 42.42, and 42.44322 by the Dutch participants. Error bars stand for ±1 standard error of the mean.*

Only "yes" responses (which indicated that the listeners performed the mapping) were included in the data analysis for the reaction time data. A linear mixed-effects model was constructed following the same steps as in Experiment 1 and 2. The model showed that TONE (42.24/42.42/42.44322) was a significant predictor of the model ($\chi^2(2) = 10.25$, $p < .01$ **). Post-hoc paired comparisons revealed: a) significantly shorter reaction time when the listeners performed the mapping in the 42.42 condition than in the 42.24 condition (Estimate = -1.34, z =

-3.10, p < .01 **); b) marginally significantly shorter reaction time in the 42.44322 condition than in the 42.24 condition (Estimate = -1.00, z = -2.30, p = .055); c) no difference between the detection rates of the 42.44322 and the 42.42 conditions (Estimate = 0.34, z = 0.80, p = .70). Figure 11 depicts the reaction time across the three conditions.

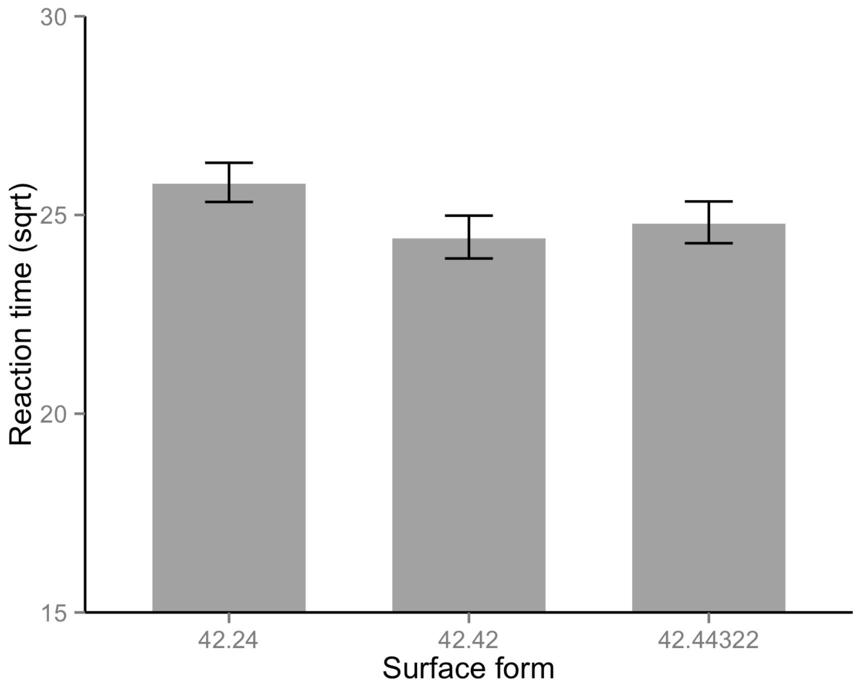

Figure 11: *Reaction time (sqrt) of the Dutch listeners' mapping in the surface forms 42.24, 42.42 and 42.44322 ("yes" responses only). Error bars stand for ±1 standard error of mean.*

The significantly higher detection rate and shorter reaction time in the surface form condition 42.42 (+ f0 disagreement, + repetition) than in the surface form condition 42.24 (- f0 disagreement, - repetition) was replicated in the current experiment, which further confirmed that the segmentation of the surface tone 42 was enhanced for the Dutch listeners in the 42.42 tone sequence.

The new condition with the surface form 42.44322 (+ f0 disagreement, -repetition) elicited a detection performance (in both detection rate and reaction time) comparable with the 42.42 condition. This persuaded us to believe that the exact repetition of the tone 42 was unlikely to be the main cause of easier tone segmentation in this study. The 42.44322 condition did not elicit higher detection rate or shorter reaction time than the 42.24 condition in a strictly statistical sense; however, the consistent evidence from the detection rate and reaction time data led us to believe that the f0 disagreement was more likely to be the major cause of enhanced segmentation by the Dutch listeners.

4.3 General discussion and conclusions

The current study investigated whether naïve non-native listeners can map surface forms of sandhied tones onto their corresponding underlying forms. For naïve non-native listeners who, by definition, lack underlying representations for the target language, the "underlying" level is presumably composed of hypothetical underlying forms that listeners deduced from surface forms by means of language non-specific articulatory knowledge such as (universal) articulatory gestures. Our study approached the non-native surface-to-underlying mapping problem by looking into a) whether assimilatory tone sandhi processes (rather than dissimilatory ones), which are motivated by (universal) knowledge of articulatory ease, allow for a transparent surface-to-underlying mapping, and b) whether this surface-to-underlying mapping is restricted only to assimilatory processes that involve gradient sound changes (rather than categorical ones) that leave residual phonetic cues of the underlying element in the surface signal.

With respect to research question a), Experiment 1 compared surface-to-underlying mappings for assimilatory and dissimilatory tone sandhi processes that result in categorical changes and provided no evidence in detection rate or reaction time to suggest that the Dutch participants are more successful in surface-to-underlying tone mapping in the assimilatory condition compared to the dissimilatory condition. Hence, we found no evidence to support the language-general surface-to-underlying mapping hypothesis for assimilatory processes.

Regarding research question b), Experiment 2 adopted more gradient sound changes, and revealed the hypothesized better mapping in the assimilatory condition based on the reaction time data, accompanied by a trend of this difference in the detection rate data. These results can be interpreted as evidence supporting the gradient-based language-general mapping hypothesis for assimilatory processes, which attaches importance to residual phonetic traces of the underlying sound in the surface form, and is generally based on the Motor Theory (Liberman et al., 1967; Liberman & Mattingly, 1985) and the Direct Realist Theory (Fowler, 1986, 1996, 2006; Fowler & Brown, 2000), which would predict the viability of non-native mapping for articulatory changes. This is the first piece of evidence, to our knowledge, that naïve non-native listeners are able to perform more successful surface-to-underlying sound mapping for (gradient) assimilatory processes than for dissimilatory processes. This difference emerged in the Dutch listeners when the maximally distant phenomena of tone and tone sandhi were used, showing the role of language-general knowledge of articulatory gestures in the mapping. However, this role was not so robust, as it was vulnerable to a segmentation effect (in categorical dissimilatory processes), and only emerged when this segmentation effect was under control (in gradient processes).

The finding that naïve non-native listeners performed better surface-to-underlying mapping only for assimilatory processes that are gradient in nature is in accordance with findings in previous studies targeting context effects which also observed the effects only for gradient processes (Gow & Im, 2004; Mitterer, Csépe, Honbolygo, et al., 2006; Mitterer et al., 2013). Nevertheless, our results do not suggest that categorical assimilatory processes would hinder naïve non-native listeners from performing surface-to-underlying mapping. Instead, we suggest that in Experiment 1 of this study, the Dutch listeners probably failed to better perform the mapping in these categorical processes because they were distracted by enhanced tone segmentation cues in the other (dissimilatory) condition.

Another conclusion of the current set of experiments is that in categorical assimilatory and dissimilatory tone sandhi processes, tone segmentation cues may play a dominant and consistent role in naïve non-tonal language (Dutch) listeners' surface-to-underlying tone mapping task. To be more specific, we suggested that the listeners may benefit from categorical dissimilatory processes because the enhanced segmentation cue (i.e., abrupt tonal disagreement) provided by these processes allows them to better locate the surface tone in its context, which gives the surface tone a more transparent connection with the (underlying) target tone. In particular, the Dutch listeners seemed to have segmented the surface tone from the disyllabic sequence more easily in the dissimilation condition 42.42; whereas they were possibly less facilitated in segmenting the surface tone from its following tonal context in the assimilation condition 42.24. Experiment 3 examined whether it is f0 disagreement in the tone sequence (defined by dissimilatory tone processes) or the repeated tone sequence that facilitates segmentation for the Dutch participants, but the two effects remained entangled; however, converging evidence from the detection rate and the reaction time data suggests that f0 disagreement is more likely to be the main cause. The observation of f0 disagreement as a cue to suggest boundaries is in alignment with previous studies such as H.-Y. Lin and Fon (2011) and Brugos and Barnes (2014).

Dissimilation by nature implies discontinuity, and hence dissimilation may have an intrinsic relationship with facilitated segmentation, as opposed to assimilation, which implies continuity and may intrinsically lead to more difficult segmentation. Johnson (1973) already proposed that dissimilation processes function to make the boundary between stem and affix more prominent and implied the intrinsic role of segmentation in these processes. Tentatively extending this idea to non-native listening, the Dutch listeners in the current study may have been dominated by facilitated segmentation in the categorical dissimilatory condition.

The reason why Darcy et al. (2009) observed successful naïve non-native listeners' mapping even for categorical assimilatory processes (specifically in viable contexts as compared to unviable contexts) is probably that they avoided listening problems due to non-native sound inventories. That is, they tested French voicing assimilation on naïve English listeners with English sound contrasts, and tested English place assimilation on naïve French listeners with French sound contrasts, and hence avoided processing of non-native sound contrasts. For listeners from a non-tonal language background, getting a grip on tone per se may be more intrinsically difficult than dealing with segments, and a task involving lexical tone and tone sequences may have encouraged listeners to overly focus on the different boundaries in the tone sequences. In addition, the Word Detection task asks listeners to first attend to the target in isolation and then immediately retrieve the target in the context, which invites attention to segmentation.

Future studies investigating the mapping difference between categorical assimilatory and dissimilatory processes by naïve non-native listeners may need to consider this enhanced segmentation effect in categorical processes and use experimental materials and methods that minimally allow segmentation effects to play a role. Testing naïve listeners in non-native assimilatory and dissimilatory processes that involve a sound contrast which exists in their native language, along the lines of Mitterer, Csépe, Honbolygo, et al. (2006) (testing Hungarian liquid assimilation on naïve Dutch listeners using Dutch sound contrasts) and Darcy et al. (2009) (as outlined above), may be a safer option. Regarding the choice of experimental paradigm, eye-tracking deserves consideration. In eye-tracking experiments, participants are presented visually with a target (underlying sound) and distractors, and auditorily with the surface sound embedded in a sentence/phrase. This would avoid the purely auditory presentation of an underlying sound paired with its surface sound in context, and may be less likely to induce segmentation effects. However, naïve listeners need to first undergo a learning session to associate the visual target with the auditory target, in order to perform this eye-tracking task. This type of experiment is difficult to conduct in a practical sense. To sum up, using sound contrasts that are typologically familiar to listeners as stimuli seems to be more promising for future studies investigating this topic. Even though maximally naïveté of listeners may be ideal for testing the role of language-general articulatory knowledge, compromises may need to be made in order to avoid undesired influences from the processing of distantly non-native sounds.

Chapter 5

General discussion and conclusions

5.1 General research questions and hypotheses revisited

The main research goal of this dissertation was to investigate the language specificity and generality of the knowledge that allows listeners to perform surface-to-underlying sound mapping. In specific, the general research questions I pursued in this dissertation were: 1) Is the surface-to-underlying mapping dependent exclusively on language-specific knowledge, or is it also facilitated by language-general articulatory knowledge (comparing native and non-native listeners)? 2) Will the mapping difficulty be different between assimilatory and dissimilatory processes, especially for naïve non-native listeners (assimilatory vs. dissimilatory)? 3) Is the mapping further influenced by the categoricalness/gradience of assimilatory and dissimilatory processes, for naïve non-native listeners (categorical vs. gradient)?

I hypothesized that native and naïve non-native listeners are doing the surface-to-underlying mapping in different ways. For native listeners, to perform the surface-to-underlying mapping means to undo the underlying-to-surface neutralization process and recover the underlying identity from the surface form, relying on their native phonological and lexical knowledge. Contrastively, for naïve non-native listeners, who lack the phonological knowledge or lexical underlying representations of the target language, what they are hypothetically doing in a surface-to-underlying mapping task is to map the surface sound onto a hypothetical underlying sound, presumably through some language-universal articulatory knowledge.

The specificity versus generality of the knowledge to allow surface-to-underlying sound mapping can be tested through assimilatory and dissimilatory processes. Assimilatory processes are mostly thought to be grounded in articulatory ease (e.g., Grammont, 1933; Johnson, 1973). The Motor Theory (Liberman et al., 1967; Liberman & Mattingly, 1985) and the Direct Realist Theory (Fowler, 1986, 1996, 2006; Fowler & Brown, 2000) hold that acoustic events are perceived as being caused by articulatory gestures. The implicit knowledge about articulatory gestures, either generalized from native articulatory experience, or from general articulator settings, is to some extent at least, language general. As a result, I hypothesized the surface-to-underlying sound mapping in assimilatory processes more likely to be language-general (the language-general mapping hypothesis for assimilatory processes). In contrast, dissimilatory processes are standardly accounted for by the "hyper-correction" theory (Ohala, 1993) or, less commonly, by the "motor planning" theory (Frisch, 2004; Frisch et al., 2004; Garrett, 2015; Garrett & Johnson, 2012; Grammont, 1895, 1933; Tilsen, 2007). The "hyper-correction" view implies that dissimilatory processes crucially refer to native lexical underlying representations that are, by definition, unavailable for naïve non-native listeners. Hence, according to this view, the surface-to-underlying mapping in dissimilatory processes is more likely to be accessible only to native listeners, and inaccessible for naïve non-native listeners (the language-specific mapping hypothesis for dissimilatory processes). The "motor planning" view, however, does not derive clear and consistent predictions regarding the language specificity/generality of the mapping for dissimilatory processes. For this reason, I only tested the language-specific mapping hypothesis for dissimilatory processes undisputedly predicted by the "hyper-correction" view in this dissertation.

As a specification of the language-general mapping hypothesis for assimilatory processes based on articulatory gestures, the surface-to-underlying mapping may be available to non-native listeners in an assimilatory process only in case the underlying representation is to some extent acoustically recoverable from the surface (the gradient-based language-general mapping hypothesis for assimilatory processes). I hypothesized that for non-native listeners, the presence of acoustic residues of the underlying sound is crucial in the mapping from the surface to the underlying forms.

This dissertation looked into tone and tone sandhi phenomena to investigate the above hypotheses on surface-to-underlying mapping, because non-tonal language listeners are maximally distant from native tonal language listeners, and thus may be ideal for testing the effect of language-general articulatory knowledge in the mapping issue.

The next section will summarize the main findings in this dissertation in a chapter-by-chapter order. Section 5.2.1 will first present the results of Chapter 2 from a production experiment on Nanjing Mandarin. This study aimed to acoustically examine the production of two pairs of comparable assimilatory and dissimilatory tone sandhi processes in Nanjing Mandarin documented in the literature, in order to verify them and decide if any of them can be used as the basis for creating stimuli in the mapping study in Chapter 3. Section 5.2.2 will summarize the findings of Chapter 3, which is a mapping study dealing with the general research questions 1) and 2), using Nanjing Mandarin as the target tonal language, and Concept Formation as the experimental paradigm. Section 5.2.3 will summarize the findings of a second mapping study in Chapter 4, which focused on surface-to-underlying mapping by naïve non-native listeners. It continued to investigate the general research question 2), using a cognitively less challenging task - Word Detection; meanwhile, it investigated the general research question 3).

5.2 Summary of main findings

5.2.1 A pair of comparable assimilatory and dissimilatory tone sandhi processes in Nanjing Mandarin

The production study in Chapter 2 tested two pairs of comparable assimilatory and dissimilatory tone sandhi processes in Nanjing Mandarin that are documented in the literature. Sandhi Pair 1 involves an underlying-surface alternation between Nanjing Tone 1 and Tone 4, and is composed of assimilatory Sandhi 1 (T1.T1 → T4.T1) and dissimilatory Sandhi 2 (T4.T5 → T1.T5); Sandhi Pair 2 involves a change between Nanjing Tone 3 and Tone 2, and is composed of assimilatory Sandhi 3 (T3.T1 → T2.T1) and dissimilatory Sandhi 4 (T3.T3 → T2.T3). The study compared the produced sandhied tone patterns by native Nanjing Mandarin speakers against the target surface forms in the literature, and identified Sandhi 1 and Sandhi 4 as categorical processes, and Sandhi 2 and Sandhi 3 as non-categorical processes. The majority of the speakers applied Sandhi 2, as opposed to more non-applications in Sandhi 3； also, Sandhi 2 applied more "categorically" than Sandhi 3. Consequently, Sandhi 1 and Sandhi 2 were selected as a pair of assimilatory and dissimilatory tone sandhi processes that are more comparable with each other. They are used as the basis for creating stimuli in the following surface-to-underlying tone mapping study in Nanjing Mandarin.

This study, as a stand-alone study, aimed to test 1) the articulation-ease motivation for assimilatory processes, which allows for gradient assimilatory processes; and 2) the "hyper-correction" theory versus the "motor planning" theory of the motivation for dissimilatory processes, with only the former theory strictly predicting categorical dissimilatory processes. It found that the two assimilatory tone sandhi processes differ in categoricalness/gradience, which agrees with Kiparky's prediction based on the articulation-ease motivation for assimilatory processes. It also observed that the two dissimilatory tone sandhi processes differ in categoricalness/gradience, which is incongruent with the prediction of the "hyper-correction" theory.

In addition, this study found inter-speaker variation differences in the four tone sandhi processes, and consequently proposed that inter-speaker variation should be standardly reported in future experimental studies on the production of tone sandhi.

Although evidence was found in this tone sandhi production task that disfavors the "hyper-correction" theory of dissimilatory processes, this evidence did not lead to changes regarding the hypothesis on mapping we assumed for dissimilatory processes based on this theory in the following experiments, because (as argued above in Section 5.1) this theory provides a clear prediction for the surface-to-underlying sound mapping, whereas the "motor planning" theory does not.

5.2.2 Surface-to-underlying mapping by native and non-native listeners in Nanjing Mandarin

In order to test the language-general mapping hypothesis for assimilatory processes and the language-specific mapping hypothesis for dissimilatory processes, the mapping study in Chapter 3 involved a native Nanjing Mandarin group and a completely naïve non-tonal language Dutch group. In addition, it included a Beijing Mandarin group, which is a non-native tonal language group, in order to a) single out the influence from native phonological knowledge on surface-to-underlying mapping, by comparing the two tonal-language groups; and b) to observe if any failure in surface-to-underlying mapping by the Dutch group was due to an incapability of performing a tone task, by comparing the two non-native language groups.

The results showed that the native Nanjing listener group successfully performed surface-to-underlying tone mapping in both the assimilatory and the dissimilatory Nanjing Mandarin sandhi processes. Based on their native phonological and/or lexical knowledge, they were found able to undo the neutralization of contextual tonal change and to perform the mapping from the sandhied tone to its corresponding underlying tone, as I predicted.

The naïve Dutch listeners failed to make use of the assimilatory context to construe an articulatory-based link between the surface tone and the underlying tone in this experiment, which contradicts my prediction. Instead, they performed around chance level in both surface-to-underlying mapping and surface-to-surface mapping. This group probably managed to form fragile and unstable categories/quasi-categories for the Nanjing Mandarin target tones through a short but intensive exemplar-based tone training task, but experienced extreme difficulty in maintaining these tone categories/quasi-categories and using them to perform the mapping task in the test session.

Adding the Beijing listeners as a control group (reasons argued above in this section) proved to be successful methodologically. On the one hand, they lack experience with the Nanjing phonological and lexical knowledge that is needed to undo the underlying-to-surface neutralization induced by the tonal context, and as such they failed to perform the mapping between Nanjing surface and underlying tones. The comparison between the performances of this group and the native Nanjing group make it clear that the native listeners established the mapping based on their native language-specific phonological and/or lexical knowledge. On the other hand, the Beijing listeners interpreted Nanjing Mandarin tones in terms of Beijing Mandarin tones at the surface level highly successfully. This is in contrast with the Dutch group's failure to map the Nanjing sandhied tones to Nanjing surface tones, and adds to the evidence that the Dutch group failed to perform the mapping at least partly because they failed to maintain tonal categories during the experiment and to use these categories in the mapping task.

The results of this mapping study confirmed the role of language-specific knowledge in surface-to-underlying tone mapping by native listeners. Nevertheless, the results did not deny the role of language-general articulatory knowledge in allowing naïve non-native listeners to perform the surface-to-underlying mapping. I suggest that the task employed in this study (Concept Formation) may have been too cognitively challenging for the Dutch listeners, when the listeners were trained to learn a tone category in a short period of time (around 20 minutes) and then asked to utilize this temporally learned tone category in the mapping task. Previous studies that successfully trained non-tonal language listeners to learn tone categories mostly adopted much longer training sessions (around or more than two weeks) to ensure that the categories were acquired in a reliable way (e.g., Chandrasekaran et al., 2010; Francis et al., 2008). Therefore, a relatively short experimental setting would call for another experimental paradigm to test the Dutch listeners.

5.2.3 Surface-to-underlying mapping by naïve Dutch listeners in an artificial tonal language: language-general mapping and enhanced segmentation

As a continued effort to investigate the effect of assimilatory vs. dissimilatory processes in non-native surface-to-underlying sound mapping, the mapping study in Chapter 4 adopted a cognitively less challenging task (Word Detection) to test Dutch listeners. Meanwhile, it aimed to investigate whether the hypothesized facilitated mapping in assimilatory processes might be observed in naïve non-native (Dutch) listeners when the process of contextual change is gradient, rather than categorical as in the study of Chapter 3 (the gradient-based language-general mapping hypothesis for assimilatory processes).

Experiment 1 in this study compared surface-to-underlying mappings for assimilatory and dissimilatory tone sandhi processes that both result in categorical changes and provided no evidence to suggest that the Dutch participants are more successful in surface-to-underlying tone mapping in the assimilatory condition compared to the dissimilatory condition.

Experiment 2 turned to more gradient sound changes, and revealed the hypothesized better mapping in the assimilatory condition with reaction time data and a trend of this difference with detection rate data. These results can be interpreted as evidence supporting the gradient-based language-general mapping hypothesis for assimilatory processes.

The finding that naïve non-native listeners performed better surface-to-underlying mapping only for assimilatory processes that are gradient does not necessarily suggest that categorical assimilatory processes would hinder naïve non-native listeners from performing surface-to-underlying mapping. Instead, I suggested that in Experiment 1 of this study, the Dutch listeners probably failed to show a mapping advantage for categorical assimilatory processes because they were distracted by enhanced tone segmentation cues within the same task (in the dissimilatory condition).

In categorical assimilatory and dissimilatory processes, tone segmentation cues may play a dominant and consistent role in naïve non-tonal language (Dutch) listeners' surface-to-underlying tone mapping. In Experiment 1, listeners may have perceptually benefited from categorical dissimilatory processes because the segmentation cue intrinsic in these processes allowed them to better locate the surface tone in its context, which gives the surface tone a stronger connection with the (underlying) target tone; whereas they seemed to be less facilated in segmenting the surface tone from the tonal context in the categorical assimilatory 42.24 condition. The dissimilatory process adopted in Experiment 1 also featured tone repetition in the surface tone sequence. Experiment 3 examined whether it is abrupt f0 disagreement (defined by categorical dissimilatory processes) in the tone sequence or the repeated tone sequence that enhanced segmentation for the Dutch participants, but the two effects remained entangled; however, the abrupt f0 disagreement was more likely to be the main cause.

In sum, the three mapping studies reported in this chapter provided some empirical evidence supporting the gradient-based language-general mapping hypothesis for assimilatory processes. However, these studies did not yield evidence suggesting that categorical assimilatory processes would hinder naïve non-native listeners from performing the mapping. Instead the results suggested that language-general segmentation cues, intrinsic in these (categorical) dissimilatory processes, may play a role in the tone mapping task.

5.3 General discussion

This dissertation began with the general research goal of investigating language specificity and generality in the mapping between surface and underlying sounds. It approached this research goal through a focus on assimilatory and dissimilatory processes, the former bridging underlying and surface sounds for listeners via language-universal articulatory ease, while the latter referring to listeners' language-specific underlying representations. Accordingly, it proposed the language-general mapping hypothesis for assimilatory processes (and a more specified gradient-based language-general mapping hypothesis for assimilatory processes) next to the language-specific mapping hypothesis for dissimilatory processes. It selected tone and tone sandhi phenomena to study, with Nanjing Mandarin listeners as native listeners and Dutch listeners as completely naïve non-native listeners, and has provided several findings regarding native and non-native surface-to-underlying mapping to be discussed below.

5.3.1 Native surface-to-underlying mapping

As predicted, native listeners were observed to be able to successfully perform the mapping from surface tones to their underlying tones in both assimilatory and dissimilatory processes. This finding added to the existing bulk of evidence on native surface-to-underlying mapping in assimilatory processes such as vowel nasalization (Beddor & Krakow, 1999; Lahiri & Marslen-Wilson, 1991), coronal place assimilation (Coenen et al., 2001; Darcy et al., 2009; Gaskell & Marslen-Wilson, 1996, 1998; Gaskell & Snoeren, 2008; Gow, 2003; Mitterer & Blomert, 2003), voicing assimilation (Coenen et al., 2001; Darcy et al., 2009; Gow & Im, 2004), labial-to-velar assimilation (Mitterer et al., 2013), and tonal assimilation (Xu, 1994); as well as to the small body of mapping studies on tonal dissimilatory processes (A. Chen & Kager, 2011; A. Chen et al., 2015; Peng, 2000).

In the Concept Formation study reported in Chapter 3, native surface-to-underlying mapping occurred even though the underlying-to-surface tonal change (in both assimilatory and dissimilatory cases) was categorical, leaving no acoustic traces of the underlying tones. Furthermore, this native mapping was shown to be not entirely dependent on lexical representations, because although the study used existing Mandarin characters in the test words, it controlled for lexicality and word frequency effects since the surface disyllabic word and the underlying disyllabic word were equally familiar or non-wordlike to the listeners. These observations are compatible with findings in previous studies, and suggest that the native listeners establish surface-to-underlying tone mappings on the basis of their native phonological knowledge and/or lexical knowledge.

The detection rate of native surface-to-underlying mapping in the current dissertation was observed to be around 75%, which is significantly below perfect mapping. This observation is also in accordance with levels of performance for mapping documented in previous studies. In segmental studies, for instance, the priming effect that Coenen et al. (2001) found in native German listeners on German place and voice assimilations was smaller in effect size for the assimilated words in appropriate contexts than for unchanged words; Darcy et al. (2009) found that French listeners detected the target underlying word of a native voicing assimilation rule in viable context at a rate of 65% (chance level: 50%), remarkably below the 96% detection rate found in the no-change control condition. Regarding tonal studies, native listeners' identification accuracy of a changed tone in its original context found by Xu (1994) was higher: ranging from 81% to 99.7% correct (chance level:

25%), but the ceiling 97% accuracy was observed when the coarticulated tone added very little variation to the underlying tone identity.

The observed 75% detection rate is explained by two sets of results in the current native surface-to-underlying mapping study (Concept Formation study, Chapter 3): 1) surface-to-underlying mapping; 2) surface-to-surface mapping. This mapping study improved on the previous native mapping studies in that it measured both sets of mapping results in one study. It was observed that the native listeners also performed the mapping between the sandhied tones and the surface tones based on the surface value of the sandhied tones. Hence, the surface-to-underlying mapping and the surface-to-surface mapping coexist and are competing with each other, which is likely the reason why the surface-to-underlying performance levels fall significantly below 100%, both in this dissertation and in the previous studies.

The native Nanjing Mandarin listeners, even though some of them may be producers of a gradient tone sandhi process (recall that Sandhi 2 proved to be gradient in Chapter 2), successfully mapped the categorically sandhied tones to their underlying tones (Sandhi 2 was used in Chapter 3 in its categorical form). This is probably because all Nanjing Mandarin listeners have been extensively exposed to both gradient and categorical forms of this sandhi process in daily life, and hence developed successful mapping for all forms of this sandhi processes.

5.3.2 Non-native surface-to-underlying mapping

The findings regarding surface-to-underlying mapping by naïve Dutch listeners in this dissertation varied depending on the categoricalness/gradience of the assimilatory and dissimilatory processes.

Dutch listeners only demonstrated facilitated surface-to-underlying mapping in gradient assimilatory processes as compared to gradient dissimilatory processes in this dissertation (Chapter 4). This observation supports the gradient-based language-general mapping hypothesis for assimilatory processes, which attaches importance to fine-grained phonetic traces of the underlying sound in the surface form, and is generally based on the Motor Theory (Liberman et al., 1967;

Liberman & Mattingly, 1985) and Direct Realist Theory (Fowler, 1986, 1996, 2006; Fowler & Brown, 2000) of speech perception, both of which predict the viability of non-native mapping for articulatory changes. This is the first piece of evidence, to my knowledge, that naïve non-native listeners are able to perform more successful surface-to-underlying sound mapping for (gradient) assimilatory processes than for dissimilatory processes. This observed difference in mapping difficulty emerged in the Dutch listeners when the maximally distant speech phenomena of tone and tone sandhi were used, showing the role of language-general knowledge of articulatory gestures in the mapping. However, this role was not robust, as it was vulnerable to a segmentation effect (in categorical dissimilatory processes), and only emerged when this segmentation effect was cleared (in gradient processes).

Reaction time, rather than detection rate, provided evidence for the facilitated surface-to-underlying mapping by Dutch listeners in this dissertation. Though Dutch listeners were observed to show comparable mapping rates in the gradient assimilatory and dissimilatory conditions, they responded faster when performing the mapping in the assimilatory condition, showing that they were more certain about the mapping in this condition. Interestingly, in a phoneme monitoring study targeting context effects by naïve listeners (Gow & Im, 2004), facilitated monitoring of the underlying sound in the viable context was also only observed with reaction time data, but not with the accuracy data that they collected.

The mapping rate in Gow and Im (2004) was higher (around 85%) than in the current study (around 55%). This difference may be due to their use of Hungarian sounds to test American English listeners, which are presumably more familiar to these listeners than lexical tones are to the Dutch listeners. The use of tone, and tone sandhi phenomena in particular, may have caused the Dutch listeners to be less sure about their responses in general, and hence they showed an overall decrease in the accuracy of their mapping in both the assimilatory and dissimilatory conditions in this study. In order to subject the (gradient-based) language-general mapping hypothesis for assimilatory processes to a strong test, maximally distant speech phenomena were used to test naïve non-native listeners in this study. Arguably, a price was paid for this decision in that this large cross-language distance caused an overall drop in the mapping rate by the non-native listeners.

When presented with categorical tone sandhi processes, Dutch listeners were not observed in this dissertation to succeed in a surface-to-underlying mapping advantage for assimilatory (compared to dissimilatory) processes. Most previous studies targeting contextual effects similarly only observed effects for gradient processes, but not for categorical processes (Gow & Im, 2004; Mitterer, Csépe, Honbolygo, et al., 2006; Mitterer et al., 2013). Nevertheless, the current results cannot be interpreted as suggesting that the categoricalness of assimilatory processes would generally prohibit surface-to-underlying mapping in naïve non-native listeners. Instead, I suggest that the Dutch listeners probably failed to show a mapping advantage for categorical assimilatory (over dissimilatory) processes because they were distracted by enhancement of tone segmentation cues within the same task (the dissimilatory condition). That is, the naïve non-native Dutch listeners seemed to have segmented the surface tone more successfully in the dissimilatory condition, where the surface tone jumped abruptly away from its tonal context in pitch (f0), as evidenced by the finding that the Dutch listeners mapped the surface tones to underlying tones more successfully in this condition. This is compared to the assimilatory condition in which the surface tone formed a continuous sequence with the tonal context in f0. However, when the change from the underlying to surface tone turned gradient (Experiment 2), making the f0 distance between the surface tone and the tonal context equal for assimilatory and dissimilatory processes, the segmentation advantage of the dissimilatory condition disappeared, and thus the hypothesized better mapping in the assimilatory condition emerged.

The role of dissimilation as a potential segmentation cue was proposed by Johnson (1973), who observed the prevalence of dissimilation phenomena across concatenated morphemes cross-linguistically, and suggested that dissimilation processes apply to enhance stem boundaries. This view implies that dissimilation intrinsically introduces segmentation cues in a language-general way. This segmentation cue resides in discontinuity and hence dissimilation may have an intrinsic relationship with easier segmentation, as opposed to assimilation, which implies continuity and may intrinsically lead to more difficult segmentation.

Evidence from previous studies also suggests that f0 disagreement can be used to suggest boundaries. For example, H.-Y. Lin and Fon (2011) observed an effect of pitch reset on the detectability of discourse boundaries and on subjects' ranking of boundary hierarchy; Brugos and Barnes (2014) found that silent intervals bounded by tokens of widely differing pitch are heard as longer than those bounded by tokens closer in pitch.

The reason why Darcy et al. (2009) observed successful naïve non-native listeners' mapping for categorical assimilatory processes is probably that they tested French voicing assimilation on naïve English listeners with English sound contrasts, and tested English place assimilation on naïve French listeners with French sound contrasts, and hence avoided processing of non-native sound contrasts. For listeners from a non-tonal language background, processing tone sequences per se may be more intrinsically difficult than processing segment sequences, and a task involving lexical tone and tone sequences may have more easily lead them to overly focus on the different boundaries in the tone sequences.

Inviting naïve Dutch listeners to perform the surface-to-underlying mapping in a cognitively challenging task while not providing them with sufficient training may have caused the listeners to fail in the mapping overall. In the Concept Formation experiment (Chapter 3), the Dutch listeners performed around chance level not only in surface-to-underlying mapping but also in pure surface-level tone mapping. This suggests that they encountered difficulties in maintaining newly formed tonal (quasi-) categories that had been introduced during a short-period training and thus had trouble relying on these (quasi-) categories during the mapping task. Indeed, previous studies that successfully trained non-tonal language listeners to learn tone categories adopted relatively longer training time (Chandrasekaran et al., 2010; Francis et al., 2008).

5.3.3 The Concept Formation experiment vs. the Word Detection experiment: strengths and weaknesses

The Concept Formation experiment (in Chapter 3) and the Word Detection experiment (in Chapter 4) displayed methodological advantages and disvantages.

5.3.3.1 The Concept Formation experiment

The Concept Formation paradigm invites listeners to first make use of a sound category (if they already possess the category) or form a sound category (if they do not possess the category yet) in a training session, and then to utilize the category in the mapping test as an underlying category of the surface sound (embedded in context). This paradigm has proven to be a successful tool to test listeners on native phonological processes: English place assimilation in Jaeger (1986), Mandarin tonal dissimilation in Peng (2000), and Nanjing Mandarin tonal assimilation and dissimilation in the current dissertation (Chapter 3). The inclusion of a training session in this paradigm theoretically made it possible to test non-native listeners as well. It permitted me to compare the three different native and non-native language groups under identical experimental conditions, and to assess the role of language-specific effects. However, this task of carrying the newly formed categories to the test in order to perform mapping proved to be too challenging for the naïve Dutch listeners when dealing with non-native tone mappings, as it failed to elicit any reliable mapping results of this group.

In the current Concept Formation experiment (in Chapter 3), I used the training session to train half of the participants on the underlying tone and the other half on the surface tone of the tone sandhi process. This allowed me to measure to what extent listeners 1) mapped the sandhied tone to the underlying tone, as well as 2) mapped the sandhied tone to the surface tone. This comparison between the two mappings made it possible for me to a) discover the co-existing and competing mapping strategies of the native Nanjing listeners; and b) conclude that the naïve Dutch listeners failed to maintain tonal categories during the experiment as shown by their failure to use these categories in the surface mapping task.

5.3.3.2 The Word Detection experiment

The Word Detection paradigm works for both native and naïve non-native listeners. It presents the isolated underlying target sound paired with the surface sound embedded in context to the listeners, and thus assumes no category formation by the listeners. This paradigm imposes much less cognitive pressure on naïve listeners (compared to the Concept Formation experiment), and allows naïve non-native listeners to perform the mapping on the basis of language-general articulatory knowledge. For example, Darcy et al. (2009) observed in this paradigm that naïve French listeners succeeded in mapping surface stop consonants with different places of articulation to underlying alveolar consonants in an English place assimilation rule, and naïve English listeners mapped surface voiced consonants to underlying voiceless consonants in a French voicing assimilation rule. This paradigm also proved to be useful in bringing out Dutch listeners' surface-to-underlying mapping preference for gradient assimilatory (compared to gradient dissimilatory) tone sandhi processes (Experiment 2 of Chapter 4).

In the Word Detection paradigm, the way in which stimuli are presented – underlying sound (presented as a phonetic sound) in isolation paired with surface sound in context – may invite naïve listeners to attend to the segmentation of the surface sound from its context when performing the mapping, depending on the presence of apparent segmentation cues to rely on. For example, in Experiments 1 and 3 of Chapter 4, the juxtaposition of the categorical assimilatory and dissimilatory conditions may have introduced a segmentation cue into the task, and the use of tone sequences may have enhanced the segmentation cue. As mentioned before, Darcy et al. (2009) did not find the mapping to be overridden by a segmentation effect in their Word Detection study, probably (as argued in Section 5.3.2) because they adopted non-native assimilatory processes that involved sound contrasts already existing in the listeners' native language, and hence avoided the processing of non-native speech sounds.

5.4 Suggestions for future studies

This dissertation provided evidence for language specificity (in the sense of listeners' reliance on the native phonology) as well as for language generality (in the sense of listeners' reliance on general knowledge of articulatory gestures) regarding surface-to-underlying sound mapping, for assimilatory and dissimilatory tone sandhi processes. In addition, the results reported in this dissertation suggest that language-general segmentation plays a role in the mapping for categorical dissimilatory processes (in juxtaposition to categorical assimilatory processes). Future studies investigating the surface-to-underlying mapping problem for naïve non-native listeners in categorical processes may need to consider this segmentation effect, and use experimental methods that minimize its role.

Three factors may have contributed to the segmentation effect observed in the categorical processes in this dissertation, namely 1) the use of assimilatory vs. dissimilatory condition; 2) the use of tone sequences; and 3) the use of the Word Detection task as the experimental paradigm. Since the assimilatory vs. dissimilatory setting is central to the research topic of surface-to-underlying mapping by naïve non-native listeners and thus cannot be avoided, this study provides evidence to suggest that future studies investigating mapping in categorical processes ought to avoid using typologically/phonologically distant speech phenomena such as tone sequences that could possibly enhance the segmentation cue. Testing naïve listeners in non-native assimilatory and dissimilatory processes that involve a sound contrast which exists in their native language, along the lines of Mitterer, Csépe, Honbolygo, et al. (2006) (testing Hungarian liquid assimilation on naïve Dutch listeners using Dutch sound contrasts) and Darcy et al. (2009) (testing French voicing assimilation on naïve English listeners with English sound contrasts, and testing English place assimilation on naïve French listeners with French sound contrasts), may be a more reliable option.

Regarding choice of experimental paradigm, future studies investigating non-native mapping in categorical processes may also need to avoid using a task (e.g., Word Detection) that invites segmentation to play a role in the mapping. For example, eye-tracking may deserve consideration. In eye-tracking experiments, participants can be presented visually with a target (the underlying sound) and distractors, and auditorily with the surface sound embedded in a sentence/phrase. Such a set-up avoids a purely auditory presentation of an underlying sound paired with its surface sound in context, and may be less likely to induce a segmentation effect. However, naïve listeners still need to be trained to associate the visual target with the auditory target in order to perform an eye-tracking task. Considering the difficulties that the Dutch listeners encountered in maintaining the newly-trained sounds and using these in the Concept Formation tone mapping task, an eye-tracking experiment may expose the participants to a relatively longer training session in order to make the effect of the target (underlying) tones more durable for them to be able to perform the surface-to-underlying tone mapping. However, this type of experiment will be difficult to conduct in a practical sense.

Taking together the chances and challenges that I proposed for a possible new experiment, using sound contrasts that are typologically familiar to listeners as stimuli seems to be most promising for future studies investigating surface-to-underlying mapping problem for naïve non-native listeners in categorical processes. Although maximally naïveness of listeners may be ideal for testing the role of language-general articulatory knowledge, compromises need to be made in order to avoid undesired influences from the processing of distantly non-native speech sounds. Future studies in this area may seek a balance between these two factors.

Bibiliography

Abramson, A. S. (1979). The coarticulation of tones: An acoustic study of Thai. In T. L. Thongkum, P. Kullavanijaya, V. Panupong, & K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer Phonetics and Phonology in Honour of Eugenie J. A. Henderson* (pp. 1-9). Bangkok: Chulalongkorn University Press.

Abrego-Collier, C. (2013). Liquid dissimilation as listener hypocorrection. In C. Cathcart, I-H. Chen, G. Finley, S. Kang, C. S. Sandy, & E. Stickles (Eds.), *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society* (pp. 3-17). Berkeley, CA, USA.

Andruski, J. E., & Costello, J. (2004). Using polynomial equations to model pitch contour shape in lexical tones: An example from Green Mong. *Journal of the International Phonetic Association, 34*, 125-140.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 classes [R package]. Retrieved from http://CRAN.R-Project.org/package=lme4

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language, 85*(4), 785-821.

Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *The Journal of the Acoustical Society of America, 106*(5), 2868-2887.

Begus, G. (2016). *The Phonetics of Aspirate Dissimilation: Grassmann's Law in Georgian*. Paper presented at the South Caucasian Chalk Circle, Paris.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J.

Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 289-304). Dordrecht: Kluwer Academic.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167-224). Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171-204). Timonium, MD: York Press.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org

Boll-Avetisyan, N., & Kager, R. (2014). OCP-PLACE in speech segmentation. *Language and Speech, 57*(3), 394-421.

Brugos, A., & Barnes, J. (2014). Effects of dynamic pitch and relative scaling on the perception of duration and prosodic grouping in American English. In N.Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 388-392). Dublin, Ireland.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America, 128*(1), 456-465.

Chang, Y.-C., & Hsieh, F.-F. (2012). Tonal coarticulation in Malaysian Hokkien: A typological anomaly? *The Linguistic Review, 29*, 37-73.

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, CA: University of California Press.

Chen, A., & Kager, R. (2011). The perception of lexical tones and tone Sandhi in L2: success or failure? In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 444-447). Hong Kong.

Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of Mandarin tone sandhi. *Language Sciences, 48*, 62-69.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.

Cheng, C.-C. (1973). *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.

Coenen, E., Zwitserlood, P., & Bölte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes, 16*(5/6), 535-564.

Cutler, A. (1998). The recognition of spoken words with variable representations *Proceedings of the ESCA Workshop on Sound Patterns of Spontaneous Speech* (pp. 83-92). La Baume-les-Aix, France.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology, 19*(5), 381-385.

Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. In F. Kügler, C. Féry, & R. Van de Vijver (Eds.), *Variation and Gradience in Phonetics and Phonology* (pp. 265-310). Berlin: Mouton De Gruyter.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*, 113-133.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*, 3-28.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America, 99*(3), 1730-1741.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics, 68*(2), 161-177.

Fowler, C. A., & Brown, J. M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics, 62*(1), 21-32.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*(2), 268-294.

Frisch, S. A. (2004). Language processing and segmental OCP effects. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 346-371). Cambridge: Cambridge University Press.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory, 22*(1), 179-228.

Garrett, A. (2015). Sound change. In C. Bowern & B. Evans (Eds.), *The Routledge handbook of historical linguistics* (pp. 227-248). London: Routledge.

Garrett, A., & Johnson, K. (2012). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 51-97). Oxford: Oxford University Press.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*(1), 144-158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 380-396.

Gaskell, M. G., & Snoeren, N. D. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance, 34*(6), 1632-1647.

Gilichinskaya, Y. D., & Strange, W. (2010). Perceptual assimilation of American English vowels by inexperienced Russian listeners. *The Journal of the Acoustical Society of America, 128*(2), EL80-EL85.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics, 65*(4), 575-590.

Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51*, 279-296.

Grammont, M. (1895). *La dissimilation consonantique dans les langues indo-européennes et dans les langues romanes*. Dijon: Imprimerie Darantière.

Grammont, M. (1933). *Traité de phonétique*. Paris: Delagrave.

Gu, W. (2015). Tone, intonation, and emphatic stress in L2 Mandarin speech by English and Cantonese learners. In the Scottish Consortium for ICPhS

2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK.

Hall, T. A. (2009). Liquid dissimilation in Bavarian German. *Journal of Germanic Linguistics, 21*(1), 1-36.

Hammarberg, R. (1976). The metaphysics of coarticulation. *Journal of Phonetics, 4*, 353-363.

Han, M. S., & Kim, K.-O. (1974). Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics, 2*, 223-232.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346-363.

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica, 30*(3), 129-148.

Ingram, J. C., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics, 25*, 343-370.

Jaeger, J. J. (1986). Concept formation as a tool for linguistic research. In J. J. Ohala & J. Jaeger (Eds.), *Experimental Phonology* (pp. 211-237). New York: Academic Press.

Jatteau, A., & Hejná, M. (2016). Dissimilation can be gradient: Evidence from Aberystwyth English. *Papers in Historical Phonology, 1*, 359-386.

Jatteau, A., & Hejná, M. (2018). Gradient dissimilation in Mongolian: implications for diachrony. *Papers in Historical Phonology, 3*, 28-75.

Jiangsusheng Difangzhi Bianzuan Weiyuanhui (Editorial Committee of Jiangsu Province Chorography). (1998). *Jiangsu Shengzhi Fangyanzhi (Jiangsu Province Dialect Records)*. Nanjing: Nanjing University Press.

Johnson, L. (1973). Dissimilation as a natural process in phonology. *Stanford Occasional Papers in Linguistics, 3*, 45-56.

Jun, J. (2004). Place assimilation. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 58-86). Cambridge: Cambridge University Press.

Kim-Renaud, Y.-K. (1974). *Korean consonantal phonology.* (Doctoral dissertation), University of Hawaii, Honolulu, Hawaii, USA.

Kiparsky, P. (1995). The phonological basis of sound change. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 640-670). Oxford: Basil Blackwell.

Koster, C. J. (1987). *Word recognition in foreign and native language: Effects of context and assimilation.* (Doctoral dissertation), Utrecht University, Dordrecht, The Netherlands.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38*(3), 245-294.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 637-675). New York: Mouton de Gruyter.

Li, Q., & Chen, Y. (2016). An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics, 54*, 123-150.

Li, X., & Kager, R. (2018a). An acoustic comparison between two pairs of assimilatory and dissimialtory tone sandhi processes in categoricalness/gradience. *Proceedings of the 11th International Symposium on Chinese Spoken Language Processing (in print)*. Taipei, Taiwan.

Li, X., & Kager, R. (2018b). A segmentation effect in Dutch listeners' surface-to-underlying mapping of tones. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of the 9th International Conference on Speech Prosody* (pp. 143-147). Poznan, Poland.

Li, X., Kager, R., & Gu, W. (2016). Surface vs. underlying listening strategies for cross-language listeners in the perception of sandhied tones in the Nanjing dialect. In C. DiCanio, J. Malins, J. Good, K. Michelson, J. Jaeger, & H. Keily (Eds.), *Proceedings of the 5th International Symposium on Tone Aspects of Languages* (pp. 33-37). Buffalo, NY, USA.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1-36.

Lin, H.-Y., & Fon, J. (2011). The role of pitch reset in perception at discourse boundaries. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1242-1245). Hong Kong.

Lin, M., & Yan, J. (1995). A perceptual study on the domain of tones in Standard Chinese. *Chinese Journal of Acoustics, 14*(4), 350-357.

Liu, D. (Ed.). (1995). *Nanjing Fangyan Cidian (A dictionary of the Nanjing dialect)*. Nanjing: Jiangsu Education Publishing House.

Ma, Q. (2009). Nanjing Fangyan Liangzizu Liandu Biandiao de Youxuanlun Fenxi (An OT solution to bisyllabic tone sandhi in Nanjing dialect). *Yuyan Yanjiu (Studies in Language and Linguistics), 29*(1), 27-32.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Mei, T. l. (1977). Tones and tone sandhi in 16th century Mandarin. *Journal of Chinese Linguistics*, 237-260.

Meyers, S. (1997). OCP effects in Optimality Theory. *Natural Language & Linguistic Theory, 15*(4), 847-892.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton: CRC Press.

Mitterer, H. (2003). *Understanding "gardem bench": Studies on the perception of assimilated word forms.* (Doctoral dissertation), Universiteit Maastricht, Maastricht, The Netherlands.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics, 65*(6), 956-969.

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology, 59*(8), 1395-1424.

Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science, 30*(3), 451-479.

Mitterer, H., Kim, S., & Cho, T. (2013). Compensation for complete assimilation in speech perception: The case of Korean labial-to-velar assimilation. *Journal of Memory and Language, 69*(1), 59-83.

Monaghan, P., & Zuidema, W. (2015). General purpose cognitive processing constraints and phonotactic properties of the vocabulary. In the Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the Workshop on the Evolution of Phonetic Capabilities: Causes, Constraints and Consequences. Satellite Event of the 18th International Congress of Phonetic Sciences*. Glasgow, UK.

Myers, J., & Tsay, J. (2003). Investigating the phonetics of Mandarin tone sandhi. *Taiwan Journal of Linguistics, 1*(1), 29-68.

Myers, J., & Tsay, J. (2008). Neutralization in Taiwan Southern Min tone sandhi. In Y.-C. E. Hsiao, Hsu, H.-C., Wee, L.-H., Ho, D.-A. (Ed.), *Interfaces in Chinese phonology: Festschrift in honor of Matthew Y. Chen on his 70th birthday* (pp. 47-78). Taiwan: Academia Sinica.

Nixon, J. S., Chen, Y., & Schiller, N. O. (2015). Multi-level processing of phonetic variants in speech production and visual word processing: evidence from Mandarin lexical tones. Language, *Cognition and Neuroscience*, 30(5), 491-505.

Odden, D. (1994). Adjacency parameters in phonology. *Language, 70*(2), 289-330.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior* (pp. 178-203). Chicago: Chicago Linguistic Society.

Ohala, J. J. (1987). Experimental phonology. In I. Kwon, H. Pritchett, & J. Spence (Eds.), *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society* (pp. 207-222). Berkeley, CA, USA.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237-278). London: Longman.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604-607.

Peng, S.-H. (1996). *Phonetic implementation and perception of place coarticulation and tone sandhi.* (Doctoral dissertation), Ohio State University, Columbus, OH, USA. Retrieved from https://etd.ohiolink.edu/

Peng, S.-H. (2000). Lexical versus 'phonological' representations of Mandarin sandhi tones. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 152-167). Cambridge: Cambridge University Press.

Qiu, C. (2012). Tone sandhi patterns of Zhenjiang dialect. In Q. Ma, H. Ding, & D. Hirst (Eds.), *Proceedings of the 6th International Conference of Speech Prosody* (pp. 354-357). Shanghai, China.

R Core Team. (2014). R: A language and environment for statistical computing (Version 3.1.0) [Computer program]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Song, Y. (2006). *Nanjing Fangyan Shengdiao Shiyan Yanjiu (Experimental study on tones of Nanjing dialect).* (Master's thesis), Nanjing Normal University, Nanjing, China.

Speer, S. R., Shih, C.-L., & Slowiaczek, M. L. (1989). Prosodic structure in language understanding: Evidence from tone sandhi in Mandarin. *Language and Speech, 32*(4), 337-354.

Svantesson, J.-O., & Karlsson, A. (2012). Preaspiration in modern and old Mongolian. *Suomalais-Ugrilaisen Seuran Toimituksia, 264*, 453-464.

Svantesson, J.-O., Tsendina, A., Karlsson, A. M., & Franzén, V. (2005). *The phonology of Mongolian*. Oxford: Oxford University Press.

Tilsen, S. (2007). Vowel-to-vowel coarticulation and dissimilation in phonemic-response priming. *UC Berkeley PhonLab Annual Report, 3*(3), 416-458.

Tilsen, S. (2013). Inhibitory mechanisms in speech planning maintain and maximize contrast. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 112-127). Oxford: Oxford University Press.

Veenker, T. J. G. (2017). The Zep experiment control application [Computer software]. Utrecht. Retrieved from http://www.beexy.org/zep

Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance, 36*(4), 1005-1015.

Wang, W. S., & Li, K. P. (1967). Tone 3 in Pekinese. *Journal of Speech, Language, and Hearing Research, 10*(3), 629-636.

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America, 95*(4), 2240-2253.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica, 55*(4), 179-203.

Yip, M. (1980). *The tonal phonology of Chinese.* (Doctoral dissertation), Massachusetts Institute of Technology, Cambridge, MA, USA. Retrieved from https://dspace.mit.edu/handle/1721.1/15971

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Yuan, J., & Chen, Y. (2014). 3rd tone sandhi in standard Chinese: A corpus approach. *Journal of Chinese Linguistics, 42*(1), 218-237.

Zeng, B., & Mattys, S. L. (2016). Separability of tones and rhymes in Chinese speech perception: Evidence from perceptual migrations. *Language and Speech, 60*(4), 562-570.

Zhang, J. (2014). *A sociophonetic study on tonal variation of the Wúxi and Shànghai dialects.* (Doctoral dissertation), Utrecht University, Utrecht, The Netherlands. Retrieved from https://dspace.library.uu.nl/handle/1874/300806

Zhang, J., & Liu, J. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica, 68*, 161-191.

Zhang, J., & Meng, Y. (2016). Structure-dependent tone sandhi in real and nonce disyllables in Shanghai Wu. *Journal of Phonetics, 54*, 169-201.

Appendices

Appendix A: Word list used in the Nanjing Mandarin tone sandhi production experiment

Words of T1, T1.Tx , and Tx.T1 ($_X$ =T1, T2, T3, T4, T5)

	Tone(s)	Words in Nanjing Mandarin (Chinese characters, Pinyin and gloss)			
T1	T1	花 *hua* 'flower'	天 *tian* 'sky'	山 *shan* 'mountain'	心 *xin* 'heart'
T1. Tx	T1.T1	花生 *hua sheng* 'peanut'	天窗 *tian chuang* 'skylight'	山峰 *shan feng* 'mountin'	心慌 *xin huang* 'flustered'
	T1.T2	花钱 *hua qian* 'to spend money'	天凉 *tian liang* 'cool weather'	山头 *shan tou* 'mountain top'	心情 *xin qing* 'mood'
	T1.T3	花粉 *hua fen* 'pollen'	天冷 *tian leng* 'cold weather'	山谷 *shan gu* 'valley'	心口 *xin kou* 'heart'
	T1.T4	花轿 *hua jiao* 'bridal sedan chair'	天亮 *tian liang* 'dawn'	山芋 *shan yu* 'sweet potato'	心算 *xin suan* 'mental arithmetic'

	T1.T5	花木 *hua mu* ‘flowers and trees’	天色 *tian se* ‘colour of the sky’	山药 *shan yao* ‘Chinese yam’	心急 *xin ji* ‘impatient’
Tx. T1	T1.T1	鲜花 *xian hua* ‘fresh flower’	阴天 *yin tian* ‘cloudy day’	金山 *jin shan* ‘golden hill’	空心 *kong xin* ‘hollow’
	T2.T1	麻花 *ma hua* ‘dough twist’	晴天 *qing tian* ‘sunny day’	爬山 *pa shan* ‘to climb mountain’	红心 *hong xin* ‘red heart’
	T3.T1	小花 *xiao hua* ‘small flower’	每天 *mei tian* ‘every day’	小山 *xiao shan* ‘hill’	忍心 *ren xin* ‘to have the heart to do’
	T4.T1	桂花 *gui hua* ‘sweet osmanthus’	变天 *bian tian* ‘change of weather’	大山 *da shan* ‘mountain’	背心 *bei xin* ‘vest’
	T5.T1	菊花 *ju hua* ‘chrysanthemum’	伏天 *fu tian* ‘dog days’	雪山 *xue shan* ‘snow mountain’	恶心 *er xin* ‘to feel sick’

Words of T2, T2.Tx , and Tx.T2 ($_X$ =T1, T2, T3, T4, T5)

	Tone(s)	Words in Nanjing Mandarin (Chinese characters, Pinyin and gloss)			
T2	T2	皮 *pi* ‘skin/leather’	头 *tou* ‘head’	平 *ping* ‘level’	年 *nian* ‘year’
T2. Tx	T2.T1	皮包 *pi bao*	头巾 *tou jin*	平菇 *ping gu*	年糕 *nian gao*

		'leather bag'	'kerchief'	'oyster mushroom'	'rice cake'
	T2.T2	皮鞋 *pi xie* 'leather shoes'	头皮 *tou pi* 'scalp'	平头 *ping tou* 'crewcut'	年头 *nian tou* 'year'
	T2.T3	皮影 *pi ying* 'shadow puppet'	头痒 *tou yang* 'itchy scalp'	平顶 *ping ding* 'flat top'	年尾 *nian wei* 'end of year'
	T2.T4	皮蛋 *pi dan* 'preserved egg'	头痛 *tou tong* 'headache'	平地 *ping di* 'flat ground'	年代 *nian dai* 'generation'
	T2.T5	皮尺 *pi chi* 'rape measure'	头发 *tou fa* 'hair'	平局 *ping ju* 'draw'	年历 *nian li* 'calendar'
Tx. T2	T1.T2	真皮 *zhen pi* 'genuine leather'	先头 *xian tou* 'previously'	天平 *tian ping* 'scale'	新年 *xin nian* 'new year'
	T2.T2	牛皮 *niu pi* 'cow hide'	抬头 *tai tou* 'to raise head'	荧屏 *ying ping* 'screen'	明年 *ming nian* 'next year'
	T3.T2	眼皮 *yan pi* 'eye lid'	口头 *kou tou* 'oral'	水平 *shui ping* 'level'	小年 *xiao nian* 'little New Year (a week before the lunar New Year)'
	T4.T2	肉皮 *rou pi* 'pig skin'	碰头 *peng tou* 'to meet'	太平 *tai ping* 'peaceful'	拜年 *bai nian* 'to pay a new year call'
	T5.T2	脚皮 *jiao pi* 'foot skin'	出头 *chu tou* 'to stand out'	六瓶 *liu ping* 'six bottles'	学年 *xue nian* 'academic year'

Words of T3, T3.Tx , and Tx.T3 ($_X$ =T1, T2, T3, T4, T5)

	Tone(s)	Words in Nanjing Mandarin (Chinese characters, Pinyin and gloss)			
T3	T3	水 *shui* 'water'	小 *xiao* 'small'	雨 *yu* 'rain'	手 *shou* 'hand'
T3. Tx	T3.T1	水沟 *shui gou* 'drain'	小江 *xiao jiang* 'small river'	雨天 *yu tian* 'rainy day'	手心 *shou xin* 'palm'
	T3.T2	水瓶 *shui ping* 'water bottle'	小河 *xiao he* 'small river'	雨鞋 *yu xie* 'rain boots'	手头 *shou tou* 'at hand'
	T3.T3	水饺 *shui jiao* 'boiled dumpling'	小雨 *xiao yu* 'light rain'	雨伞 *yu san* 'umbrella'	手掌 *shou zhang* 'palm'
	T3.T4	水面 *shui mian* 'water surface'	小象 *xiao xiang* 'calf elephant'	语病 *yu bing* 'grammatical mistake'	手背 *shou bei* 'back of the hand'
	T3.T5	水竹 *shui zhu* 'water bamboo'	小吃 *xiao chi* 'snack'	雨脚 *yu jiao* 'splash of rain on the ground'	手脚 *shou jiao* 'hands and feet'
Tx. T3	T1.T3	清水 *qing shui* 'clear water'	春晓 *chun xiao* 'early spring'	春雨 *chun yu* 'spring rain'	亲手 *qin shou* 'by hand'
	T2.T3	洪水 *hong shui* 'flood'	从小 *cong xiao* 'from childhood'	梅雨 *mei yu* 'plum rain'	抬手 *tai shou* 'to raise hands'
	T3.T3	口水 *kou shui*	老小 *lao xiao*	小雨 *xiao yu*	摆手 *bai shou*

		'saliva'	'grown ups and children'	'light rain'	'to wave hands'
	T4.T3	烫水 *tang shui* 'hot water'	大小 *da xiao* 'size'	阵雨 *zhen yu* 'shower'	右手 *you shou* 'right hand'
	T5.T3	喝水 *he shui* 'to drink water'	发小 *fa xiao* 'childhood friend'	落雨 *luo yu* 'to rain'	接手 *jie shou* 'to take over'

Words of T4, T4.Tx , and Tx.T4 ($_X$ =T1, T2, T3, T4, T5)

	Tone(s)	Words in Nanjing Mandarin (Chinese characters, Pinyin and gloss)			
T4	T4	地 *di* 'land'	豆 *dou* 'bean'	菜 *cai* 'vegetable'	树 *shu* 'tree'
T4. Tx	T4.T1	地瓜 *di gua* 'sweet potato'	豆浆 *dou jiang* 'soybean milk'	菜瓜 *cai gua* 'snake melon'	树根 *shu gen* 'tree root'
	T4.T2	地坪 *di ping* 'terrace'	豆芽 *dou ya* 'bean sprout'	菜头 *cai tou* 'turnip'	树皮 *shu pi* 'bark'
	T4.T3	地主 *di zhu* 'landlord'	豆米 *dou mi* 'green soy bean'	菜籽 *cai zi* 'vegetable seed'	树顶 *shu ding* 'tree top'
	T4.T4	地契 *di qi* 'land contract'	豆瓣 *dou ban* 'halves of a bean'	菜地 *cai di* 'vegetable field'	树洞 *shu dong* 'tree hole'
	T4.T5	地利 *di li*	豆角 *dou jiao*	菜叶 *cai ye*	树叶 *shu ye*

		‘topographical advantage’	‘long bean’	‘vegetable leaf’	‘leaf’
Tx. T4	T1.T4	荒地 *huang di* ‘waste land’	豌豆 *wan dou* ‘pea’	菠菜 *bo cai* ‘spinach’	栽树 *zai shu* ‘to plant trees’
	T2.T4	平地 *ping di* ‘flat ground’	黄豆 *huang dou* ‘soybean’	芹菜 *qin cai* ‘celery’	杨树 yang shu ‘aspen’
	T3.T4	本地 *ben di* ‘local’	扁豆 *bian dou* ‘lentil’	韭菜 *jiu cai* ‘garlic chives’	柳树 *liu shu* ‘willow’
	T4.T4	菜地 *cai di* ‘vegetable field’	大豆 *da dou* ‘soybean’	素菜 *su cai* ‘vegetable dish’	种树 *zhong shu* ‘to plant trees’
	T5.T4	挖地 *wa di* ‘to dig up soil’	赤豆 *chi dou* ‘red bean’	白菜 *bai cai* ‘Chinese cabbage’	挖树 *wa shu* ‘to dig up a tree’

Words of T5, T5.Tx , and Tx.T5 ($_{X}$ =T1, T2, T3, T4, T5)

	Tone(s)	Words in Nanjing Mandarin (Chinese characters, Pinyin and gloss)			
T5	T5	雪 *xue* ‘snow’	木 *mu* ‘wood’	月 *yue* ‘moon’	药 *yao* ‘medicine’
T5. Tx	T5.T1	雪花 *xue hua* ‘snowflake’	木瓜 *mu gua* ‘papaya’	月晕 *yue yun* ‘lunar aureole’	药膏 *yao gao* ‘ointment’
	T5.T2	雪球 *xue qiu*	木头 *mu tou*	月头 *yue tou*	药房 *yao fang*

		'snowball'	'wood'	'start of the month'	'pharmacy'
	T5.T3	雪水 *xue shui* 'snow water'	木板 *mu ban* 'wood board'	月尾 *yue wei* 'end of the month'	药粉 *yao fen* 'medicine in powder form'
	T5.T4	雪地 *xue di* 'snowfield'	木匠 *mu jiang* 'carpenter'	月半 *yue ban* 'the 15th day of a month'	药店 *yao dian* 'pharmacy'
	T5.T5	雪白 *xue bai* 'snow-white'	木屑 *mu xie* 'saw dust'	月历 *yue li* 'calendar'	药业 *yao ye* 'pharmaceutical industry
Tx. T5	T1.T5	春雪 *chun xue* 'spring snow'	杉木 *shan mu* 'cedar'	正月 *zheng yue* 'the first month of the lunar year'	抓药 *zhua yao* 'to fill the prescription'
	T2.T5	同学 *tong xue* 'school mate'	楠木 *nan mu* 'Phoebe zhennan'	明月 *ming yue* 'bright moon'	丸药 *wan yao* 'medicine in pill'
	T3.T5	小雪 *xiao xue* 'light snow'	朽木 *xiu mu* 'deadwood'	小月 *xiao yue* 'a 29-days month of the lunar calendar'	买药 *mai yao* 'to buy medicine'
	T4.T5	下雪 *xia xue* 'to snow'	树木 *shu mu* 'tree'	闰月 *run yue* 'leap month'	配药 *pei yao* 'to fill the prescription'
	T5.T5	滑雪 *hua xue*	积木 *ji mu*	腊月 *la yue*	喝药 *he yao*

		'to ski'	'building block'	'the twelfth month of the lunar year'	'to take medicine'

Appendix B: Random effect (intercept, linear and quadratic) estimates of the sandhied tone and the target surface tone for each speaker

The following values are rounded off to 3 decimals.

Sandhi 1: T1(syl1) in T1.T1 and T4 in T4.T1

Speaker	Tone	Intercept	Linear	Quadratic
1	T1(syl1) in T1.T1	-0.247	0.440	-0.064
	T4 in T4.T1	0.239	-0.086	-0.007
2	T1(syl1) in T1.T1	0.666	-1.141	0.164
	T4 in T4.T1	-0.520	0.393	-0.026
3	T1(syl1) in T1.T1	0.604	0.304	-0.125
	T4 in T4.T1	-0.539	0.249	0.005
4	T1(syl1) in T1.T1	-0.227	-0.089	0.042
	T4 in T4.T1	0.271	-0.128	-0.002
5	T1(syl1) in T1.T1	0.458	0.088	-0.066
	T4 in T4.T1	-0.436	0.303	-0.017
6	T1(syl1) in T1.T1	-0.088	0.152	-0.022
	T4 in T4.T1	0.097	0.131	-0.037
7	T1(syl1) in T1.T1	-0.517	0.129	0.027
	T4 in T4.T1	0.529	-0.196	-0.015
8	T1(syl1) in T1.T1	-0.532	-0.383	0.134
	T4 in T4.T1	0.560	0.399	-0.140
9	T1(syl1) in T1.T1	0.488	0.169	-0.085
	T4 in T4.T1	-0.518	-0.107	0.076
10	T1(syl1) in T1.T1	-0.390	0.016	0.037
	T4 in T4.T1	0.279	0.185	-0.067
11	T1(syl1) in T1.T1	-0.196	-2.266	0.484
	T4 in T4.T1	-0.021	0.466	-0.093

12	T1(syl1) in T1.T1	-0.327	0.324	-0.032
	T4 in T4.T1	0.357	0.112	-0.060
13	T1(syl1) in T1.T1	-0.287	0.245	-0.020
	T4 in T4.T1	0.300	0.071	-0.046
14	T1(syl1) in T1.T1	0.151	-1.217	0.233
	T4 in T4.T1	-0.177	0.267	-0.036
15	T1(syl1) in T1.T1	0.076	0.296	-0.069
	T4 in T4.T1	-0.054	0.266	-0.049
16	T1(syl1) in T1.T1	-0.129	0.293	-0.047
	T4 in T4.T1	0.126	0.232	-0.061
17	T1(syl1) in T1.T1	0.119	0.150	-0.043
	T4 in T4.T1	-0.097	-0.056	0.022
18	T1(syl1) in T1.T1	0.317	0.013	-0.036
	T4 in T4.T1	-0.334	-0.024	0.040

Sandhi 2: T4 in T4.T5 and T1 in T1.T5

Speaker	Tone	Intercept	Linear	Quadratic
1	T4 in T4.T5	0.334	-0.346	-0.137
	T1 in T1.T5	-0.336	0.344	0.145
2	T4 in T4.T5	0.254	0.572	-0.154
	T1 in T1.T5	-0.281	-0.287	0.126
3	T4 in T4.T5	0.064	-0.894	0.050
	T1 in T1.T5	-0.138	0.591	0.011
4	T4 in T4.T5	-0.170	-0.254	0.116
	T1 in T1.T5	0.162	0.046	-0.092
5	T4 in T4.T5	-0.339	1.062	0.064
	T1 in T1.T5	0.311	-0.689	-0.096
6	T4 in T4.T5	-0.062	-0.175	0.048
	T1 in T1.T5	0.094	0.225	-0.107
7	T4 in T4.T5	0.004	-0.305	0.013
	T1 in T1.T5	-0.041	0.516	-0.008
8	T4 in T4.T5	0.001	0.164	-0.015

	T1 in T1.T5	-0.015	-0.186	0.043
9	T4 in T4.T5	0.112	-0.452	-0.017
	T1 in T1.T5	-0.097	0.394	0.008
10	T4 in T4.T5	-0.097	-0.292	0.100
	T1 in T1.T5	0.035	0.452	-0.085
11	T4 in T4.T5	0.359	-0.366	-0.141
	T1 in T1.T5	-0.407	-0.376	0.282
12	T4 in T4.T5	0.160	-0.059	-0.062
	T1 in T1.T5	-0.190	0.598	0.005
13	T4 in T4.T5	-0.260	-0.194	0.145
	T1 in T1.T5	0.261	0.081	-0.124
14	T4 in T4.T5	-0.255	0.373	0.074
	T1 in T1.T5	0.203	-0.668	-0.039
15	T4 in T4.T5	0.079	0.250	-0.058
	T1 in T1.T5	-0.100	0.053	0.038
16	T4 in T4.T5	0.094	-0.516	0.012
	T1 in T1.T5	-0.121	0.684	-0.006
17	T4 in T4.T5	-0.128	0.143	0.039
	T1 in T1.T5	0.516	-0.542	-0.193
18	T4 in T4.T5	-0.149	1.289	-0.076
	T1 in T1.T5	0.142	-1.236	0.090

Sandhi 3: T3 in T3.T1 and T2 in T2.T1

Speaker	Tone	Intercept	Linear	Quadratic
1	T3 in T3.T1	0.046	1.468	0.486
	T2 in T2.T1	-0.029	-0.049	0.018
2	T3 in T3.T1	-0.347	-4.050	-1.065
	T2 in T2.T1	-0.089	2.107	0.889
3	T3 in T3.T1	-0.197	-0.905	-0.089
	T2 in T2.T1	0.175	2.184	0.590
4	T3 in T3.T1	0.071	-0.522	-0.281
	T2 in T2.T1	-0.013	-0.040	0.001

5	T3 in T3.T1	0.230	-0.790	-0.577
	T2 in T2.T1	-0.211	0.028	0.272
6	T3 in T3.T1	-0.014	0.038	0.031
	T2 in T2.T1	0.002	0.709	0.259
7	T3 in T3.T1	0.031	0.058	-0.016
	T2 in T2.T1	-0.052	-0.669	-0.183
8	T3 in T3.T1	0.116	0.849	0.170
	T2 in T2.T1	-0.048	-0.894	-0.270
9	T3 in T3.T1	0.008	-1.315	-0.495
	T2 in T2.T1	-0.069	1.235	0.542
10	T3 in T3.T1	-0.164	-0.588	-0.013
	T2 in T2.T1	0.308	1.819	0.289
11	T3 in T3.T1	0.142	1.061	0.216
	T2 in T2.T1	-0.210	-0.495	0.077
12	T3 in T3.T1	0.134	-0.337	-0.291
	T2 in T2.T1	-0.066	-1.659	-0.531
13	T3 in T3.T1	0.002	0.327	0.118
	T2 in T2.T1	0.002	0.286	0.104
14	T3 in T3.T1	-0.375	-1.192	0.025
	T2 in T2.T1	0.534	2.046	0.093
15	T3 in T3.T1	0.397	0.114	-0.450
	T2 in T2.T1	-0.296	-2.066	-0.395
16	T3 in T3.T1	-0.170	-1.103	-0.196
	T2 in T2.T1	0.265	-0.410	-0.481
17	T3 in T3.T1	0.283	2.678	0.638
	T2 in T2.T1	-0.305	0.173	0.442
18	T3 in T3.T1	-0.214	-1.482	-0.282
	T2 in T2.T1	0.124	1.384	0.357

Sandhi 4: T3(syl 1) in T3.T3 and T2 in T2.T3

Speaker	Tone	Intercept	Linear	Quadratic
1	T3(syl 1) in T3.T3	0.129	-0.052	-0.071
	T2 in T2.T3	-0.129	-0.366	0.095
2	T3(syl 1) in T3.T3	-0.107	0.050	0.058
	T2 in T2.T3	0.072	-0.059	-0.038
3	T3(syl 1) in T3.T3	0.091	-0.210	-0.04
	T2 in T2.T3	-0.059	0.372	0.013
4	T3(syl 1) in T3.T3	-0.002	-0.288	0.017
	T2 in T2.T3	0.006	-0.030	-0.002
5	T3(syl 1) in T3.T3	0.215	0.160	-0.131
	T2 in T2.T3	-0.221	-0.109	0.132
6	T3(syl 1) in T3.T3	-0.064	0.075	0.032
	T2 in T2.T3	0.046	-0.033	-0.025
7	T3(syl 1) in T3.T3	-0.218	-0.135	0.132
	T2 in T2.T3	0.182	-0.029	-0.102
8	T3(syl 1) in T3.T3	-0.047	-0.030	0.028
	T2 in T2.T3	0.099	-0.129	-0.049
9	T3(syl 1) in T3.T3	-0.168	-0.094	0.101
	T2 in T2.T3	0.125	0.302	-0.089
10	T3(syl 1) in T3.T3	-0.120	0.045	0.066
	T2 in T2.T3	0.150	0.072	-0.090
11	T3(syl 1) in T3.T3	-0.485	-0.174	0.287
	T2 in T2.T3	0.405	0.473	-0.258
12	T3(syl 1) in T3.T3	0.234	-0.085	-0.129
	T2 in T2.T3	-0.208	-0.440	0.143
13	T3(syl 1) in T3.T3	-0.075	-0.061	0.047
	T2 in T2.T3	0.100	0.394	-0.079
14	T3(syl 1) in T3.T3	0.137	-0.100	-0.072
	T2 in T2.T3	-0.106	0.327	0.042
15	T3(syl 1) in T3.T3	-0.030	1.255	-0.054
	T2 in T2.T3	0.055	-1.423	0.049
16	T3(syl 1) in T3.T3	-0.037	0.112	0.015

	T2 in T2.T3	0.069	-0.271	-0.024
17	T3(syl 1) in T3.T3	-0.309	0.471	0.150
	T2 in T2.T3	0.260	-0.119	-0.142
18	T3(syl 1) in T3.T3	-0.049	0.022	0.027
	T2 in T2.T3	0.063	0.108	-0.042

Appendix C: Tone sequences used in the Concept Formation experiment

Sandhi 1: T1 as the target tone

	Target (for training and test)		Non-target (for training and test)		Test
Syllable	Syllable 1	Syllable 2	Syllable 1	Syllable 2	
Tone	T1	T2, T3, T5	T2, T3, T5	T2, T3, (T5)	
Tone sequences	T1.T2 T1.T3 T1.T5		T2.T2, T2.T3, T2.T5 T3.T2, T3.T3, T3.T5 T5.T2, T5.T3[22]		Sandhied T1.T1

Sandhi 1: T4 as the target tone

	Target (for training and test)		Non-target (for training and test)		Test
Syllable	Syllable 1	Syllable 2	Syllable 1	Syllable 2	
Tone	T4	T2, T3, T4	T2, T3, T5	T2, T3, T4	
Tone sequences	T4.T2 T4.T3 T4.T4		T2.T2, T2.T3, T2.T4 T3.T2, T3.T3, T3.T4 T5.T2, T5.T3, T5.T4		Sandhied T1.T1

[22] T5.T5 was not included as a non-target tone sequence, because the T5 on syllable 1 was produced as a falling tone, which does not sound distinct from the target T1, according to a tone sandhi elicitation task we conducted on Nanjing Mandarin.

Sandhi 2: T4 as the target tone

	Target (for training and test)		Non-target (for training and test)		Test
Syllable	Syllable 1	Syllable 2	Syllable 1	Syllable 2	
Tone	T4	T2, T3, T4	T2, T3, T5	T2, T3, T4	
Tone sequences	T4.T2 T4.T3 T4.T4		T2.T2, T2.T3, T2.T4 T3.T2, T3.T3, T3.T4 T5.T2, T5.T3, T5.T4		Sandhied T4.T5

Sandhi 2: T1 as the target tone

	Target (for training and test)		Non-target (for training and test)		Test
Syllable	Syllable 1	Syllable 2	Syllable 1	Syllable 2	
Tone	T1	T2, T3, T4	T2, T3, T5	T2, T3, T4	
Tone sequences	T1.T2 T1.T3 T1.T4		T2.T2, T2.T3, T2.T4 T3.T2, T3.T3, T3.T4 T5.T2, T5.T3, T5.T4		Sandhied T4.T5

Appendix D: Word list used in the Concept Formation experiment

Sandhi 1: T1 as the target tone

Real-word session					
			Chinese characters	Pinyin	Gloss
Training Session	Targets (24 tokens)	T1.T2	花钱	*hua qian*	'to spend money'
			天凉	*tian liang*	'cold weather'
			山头	*shan tou*	'mountain top'
			心情	*xin qing*	'mood'
			真皮	*zhen pi*	'genuine lether'
			先头	*xian tou*	'previously'
			天平	*tian ping*	'scale'
			新年	*xin nian*	'new year'
		T1.T3	花粉	*hua fen*	'pollen'
			天冷	*tian leng*	'cold weather'
			山谷	*shan gu*	'valley'
			心口	*xin kou*	'heart'
			清水	*qing shui*	'clear water'
			春晓	*chun xiao*	'early spring'
			春雨	*chun yu*	'spring rain'
			亲手	*qin shou*	'by hand'
		T1.T5	花木	*hua mu*	'flowers and trees'
			天色	*tian se*	'colour of the sky'
			稀客	*xi ke*	'rare visitor'
			心急	*xin ji*	'impatient'
			春雪	*chun xue*	'spring snow'
			杉木	*shan mu*	'cedar'
			正月	*zheng yue*	'the first month of

					the lunar year'
			抓药	*zhua yao*	'to fill the prescription'
	Non-targets (24 tokens)	T2.T2	头皮	*tou pi*	'scalp'
			荧屏	*ying ping*	'screen'
			平行	*ping xing*	'parallel'
		T2.T3	洪水	*hong shui*	'flood'
			抬手	*tai shou*	'to raise hands'
			从小	*cong xiao*	'from childhood'
		T2.T5	同学	*tong xue*	'school mate'
			平局	*ping ju*	'draw'
			皮尺	*pi chi*	'tape measure'
		T3.T2	手头	*shou tou*	'at hand'
			眼皮	*yan pi*	'eye lid'
			水平	*shui ping*	'level'
		T3.T3	摆手	*bai shou*	'to wave hands'
			口水	*kou shui*	'saliva'
			老小	*lao xiao*	'grown-ups and children'
		T3.T5	小吃	*xiao chi*	'snack'
			买药	*mai yao*	'to buy medicine'
			水竹	*shui zhu*	'water bamboo (Phyllostachys heteroclada)'
		T5.T2	出头	*chu tou*	'to stand out'
			雪球	*xue qiu*	'snowball'
			脚皮	*jiao pi*	'foot skin'
		T5.T3	木板	*mu ban*	'wood board'
			接手	*jie shou*	'to take over'
			喝水	*he shui*	'to drink water'
Test Session	Targets (8 tokens)	T1.T2	周年	*zhou nian*	'anniversary'
			加钱	*jia qian*	'to add money'
			书房	*shu fang*	'study'

		T1.T3	车顶	*che ding*	'car roof'
			家长	*jia zhang*	'parents'
			冲水	*chong shui*	'to flush'
		T1.T5	山药	*shan yao*	'Chinese yam'
			亲切	*qin qie*	'kind and cordial'
	Non-targets (8 tokens)	T2.T2	明年	*ming nian*	'next year'
		T2.T3	平顶	*ping ding*	'flat top'
		T2.T5	头发	*tou fa*	'hair'
		T3.T2	小河	*xiao he*	'small river'
		T3.T3	水饺	*shui jiao*	'boiled dumpling'
		T3.T5	朽木	*xiu mu*	'deadwood'
		T5.T2	学年	*xue nian*	'academic year'
		T5.T3	发小	*fa xiao*	'childhood friend'
	Test tokens (16 tokens)	Sandhied T1.T1	金山	*jin*T1 *shan*T1	'gold moutain'
			进山	*jin*T4 *shan*T1	'to get into a mountain'
			精心	*jing*T1 *xin*T1	'elaborately'
			尽心	*jin*T4[23]*xin*T1	'with all one's heart'
			抛开	*pao*T1 *kai*T1	'to throw away'
			泡开	*pao*T4 *kai*T1	'to rehydrate and soften'
			心焦	*xin*T1 *jiao*T1	'worried'
			性交	*xing*T4 *jiao*T1	'sextual intercourse'
			刀边	*dao*T1 *bian*T1	'edge of the knife'
			到边	*dao*T4 *bian*T1	'to reach the edge'
			先包	*xian*T1 *bao*T1	'to wrap first'
			现包	*xian*T4 *bao*T1	'freshly made'
			梳妆	*shu*T1 *zhuang*T1	'to dress and make up'
			树桩	*shu*T4 *zhuang*T1	'stump'
			鲜花	*xian*T1 *hua*T1	'fresh flowers'
			献花	*xian*T4 *hua*T1	'to present flowers'

[23] Nanjing Mandarin does not distinguish *in* from *ing* (Liu, 1995).

			栽花	*zai*T1 *hua*T1	'to grow flowers'
			再花	*zai*T4 *hua*T1	'to spend again'
			将心	*jiang*T1 *xin*T1	'first half of the idiom "jiang xin bi xin" (to judge others' feelings by one's own)'
			匠心	*jiang*T4 *xin*T1	'ingenuity'
			商机	*shang*T1 *ji*T1	'business opportunity'
			上机	*shang*T4 *ji*T1	'to board the plane'
			川香	*chuan*T1 *xiang*T1	'Sizhuan flavor'
			串香	*chuan*T4 *xiang*T1	'to taint the flavour'
			金风	*jin*T1 *feng*T1	'autumn wind'
			进风	*jin*T4 *feng*T1	'to let wind in'
			搬箱	*ban*T1 *xiang*T1	'to carry a box'
			半箱	*ban*T4 *xiang*T1	'half a box'
			生津	*sheng*T1 *jin*T1	'to promote the secretion of saliva'
			圣经	*sheng*T4 *jing*T1	'Bible'
			搬空	*ban*T1 *kong*T1	'to empty'
			半空	*ban*T4 *kong*T1	'in the air'

Non-word session			
			Pinyin
Training Session	Targets (24 tokens)	T1.T2	*dong qian, shang ya, guo que, hua qiong, shi qie, chuan qin, quan ting, yin he*
		T1.T3	*shou xiang, cai xian, bao yan, xian xiao, pao lian, ban huo, ji bian, chun jiang*
		T1.T5	*dao ya, shu li, di yao, mao ji, zhong shua, che mu, fan jie, jian ju*
	Non-	T2.T2	*ping pi, tai qiong, tou ti*

	targets (24 tokens)	T2.T3	*hong xiu, qian bing, yuan xiang*	
		T2.T5	*ming fu, he zhu, tong qia*	
		T3.T2	*fen qie, xie qiu, xian ping*	
		T3.T3	*juan xing, xi jie, shan gan*	
		T3.T5	*tian ya, xing xi, xiao ti*	
		T5.T2	*fa lin, mu yan, ji qian*	
		T5.T3	*yao yan, ke xiao, ca xian*	
Test Session	Targets (8 tokens)	T1.T2	*jiao cao, tou fang, fen po*	
		T1.T3	*xin fen, xian bao, chao bing*	
		T1.T5	*zai jie, shen mie*	
	Non-targets (8 tokens)	T2.T2	*cong huang*	
		T2.T3	*yan chao*	
		T2.T5	*wang bie*	
		T3.T2	*qian pang*	
		T3.T3	*jie xian*	
		T3.T5	*jiao ye*	
		T5.T2	*re xing*	
		T5.T3	*la lian*	
	Test tokens (16 tokens)	Sandhied T1.T1	Pinyin	Chinese characters
			*bao*T1 *qiao*T1	包敲
			*bao*T4 *qiao*T1	抱敲
			*jiang*T1 *cai*T1	将猜
			*jiang*T4 *cai*T1	降猜
			*di*T1 *guai*T1	低乖
			*di*T4 *guai*T1	地乖
			*ji*T1 *she*T1	鸡奢
			*ji*T4 *she*T1	记奢
			*zhi*T1 *tian*T1	知天
			*zhi*T4 *tian*T1	志天
			*dao*T1 *fan*T1	刀翻
			*dao*T4 *fan*T1	到翻
			*fan*T1 *cao*T1	翻操
			*fan*T4 *cao*T1	饭操

			*suan*T1 *jian*T1	酸尖
			*suan*T4 *jian*T1	算尖
			*mao*T1 *che*T1	猫车
			*mao*T4 *che*T1	帽车
			*dao*T1 *deng*T1	刀灯
			*dao*T4 *deng*T1	到灯
			*zhuang*T1 *huang*T1	装荒
			*zhuang*T4 *huang*T1	撞荒
			*cai*T1 *jin*T1	猜金
			*cai*T4 *jin*T1	菜金
			*xun*T1 *chuan*T1	熏穿
			*xun*T4 *chuan*T1	训穿
			*jiang*T1 *chuan*T1	江川
			*jiang*T4 *chuan*T1	降穿
			*guai*T1 *xiu*T1	乖修
			*guai*T4 *xiu*T1	怪修
			*pao*T1 *bian*T1	抛边
			*pao*T4 *bian*T1	泡边

Sandhi 1: T4 as the target tone

Real-word session					
			Chinese characters	Pinyin	Gloss
Training Session	Targets (24 tokens)	T4.T2	肉皮	*rou pi*	'pig skin'
			碰头	*peng tou*	'to meet'
			太平	*tai ping*	'peaceful'
			拜年	*bai nian*	'to pay a new year call'
			地坪	*di ping*	'terrace'
			豆芽	*dou ya*	'bean sprout'
			菜头	*cai tou*	'turnip'
			树皮	*shu pi*	'bark'

		T4.T3	地主	*di zhu*	'landlord'
			豆米	*dou mi*	'green soy bean'
			菜籽	*cai zi*	'vegetable seed'
			树顶	*shu ding*	'tree top'
			烫水	*tang shui*	'hot water'
			大小	*da xiao*	'size'
			阵雨	*zhen yu*	'shower'
			右手	*you shou*	'right hand'
		T4.T4	地契	*di qi*	'land contract'
			豆瓣	*dou ban*	'halves of a bean'
			菜地	*cai di*	'vegetable field'
			树洞	*shu dong*	'tree hole'
			大豆	*da dou*	'soybean'
			素菜	*su cai*	'vegetable dish'
			种树	*zhong shu*	'to plant trees'
			教授	*jiao shou*	'professor'
	Non-targets (24 tokens)	T2.T2	头皮	*tou pi*	'scalp'
			平行	*pi xing*	'parrallel'
		T2.T3	洪水	*hong shui*	'flood'
			抬手	*tai shou*	'to raise hands'
			从小	*cong xiao*	'from childhood'
		T2.T4	皮蛋	*pi dan*	'preserved egg'
			头痛	*tou tong*	'headache'
			黄豆	*huang dou*	'soybean'
		T3.T2	手头	*shou tou*	'at hand'
			眼皮	*yan pi*	'eyelid'
			水平	*shui ping*	'level'
		T3.T3	口水	*kou shui*	'saliva'
			手掌	*shou zhang*	'palm'
		T3.T4	扁豆	*bian dou*	'lentil'
			小象	*xiao xiang*	'calf elephant'
			本地	*ben di*	'local'
		T5.T2	出头	*chu tou*	'to stand out'

			雪球	*xue qiu*	‘snowball’
			脚皮	*jiao pi*	‘foot skin’
		T5.T3	木板	*mu ban*	‘wood board’
			接手	*jie shou*	‘to take over’
			喝水	*he shui*	‘to drink water’
		T5.T4	赤豆	*chi dou*	‘red bean’
			白菜	*bai cai*	‘Chinese cabbage’
Test Session	Targets (8 tokens)	T4.T2	菜油	*cai you*	‘rape oil’
			过年	*guo nian*	‘to have the Spring Festival’
			太平	*tai ping*	‘peaceful’
		T4.T3	放水	*fang shui*	‘to drain off water’
			现场	*xian chang*	‘on the spot’
			暴雨	*bao yu*	‘rainstorm’
		T4.T4	做事	*zuo shi*	‘to do things’
			庆祝	*qing zhu*	‘to celebrate’
	Non-targets (8 tokens)	T2.T2	明年	*ming nian*	‘next year’
		T2.T3	平顶	*ping ding*	‘flat top’
		T2.T4	杨树	*yang shu*	‘aspen tree’
		T3.T2	小河	*xiao he*	‘small river’
		T3.T3	水饺	*shui jiao*	‘boiled dumpling’
		T3.T4	手背	*shou bei*	‘back of the hand’
		T5.T2	学年	*xue nian*	‘academic year’
		T5.T3	发小	*fa xiao*	‘childhood friend’
	Test tokens (16 tokens)	Sandhied T1.T1	金山	*jin*T1 *shan*T1	‘gold moutain’
			进山	*jin*T4 *shan*T1	‘to get into a mountain’
			精心	*jing*T1 *xin*T1	‘elaborately’
			尽心	*jin*T4 *xin*T1	‘with all one’s heart’
			抛开	*pao*T1 *kai*T1	‘to throw away’
			泡开	*pao*T4 *kai*T1	‘to rehydrate and soften’
			心焦	*xin*T1 *jiao*T1	‘worried’

			性交	*xing*T4 *jiao*T1	'sexual intercourse'
			刀边	*dao*T1 *bian*T1	'edge of the knife'
			到边	*dao*T4 *bian*T1	'to reach the edge'
			先包	*xian*T1 *bao*T1	'to wrap first'
			现包	*xian*T4 *bao*T1	'freshly made'
			梳妆	*shu*T1 *zhuang*T1	'to dress and make up'
			树桩	*shu*T4 *zhuang*T1	'stump'
			鲜花	*xian*T1 *hua*T1	'fresh flowers'
			献花	*xian*T4 *hua*T1	'to present flowers'
			栽花	*zai*T1 *hua*T1	'to grow flowers'
			再花	*zai*T4 *hua*T1	'to spend again'
			将心	*jiang*T1 *xin*T1	'first half of the idiom *jiang xin bi xin (to judge others' feelings by one's own)*'
			匠心	*jiang*T4 *xin*T1	'ingenuity'
			商机	*shang*T1 *ji*T1	'business opportunity'
			上机	*shang*T4 *ji*T1	'to board the plane'
			川香	*chuan*T1 *xiang*T1	'Sizhuan flavor'
			串香	*chuan*T4 *xiang*T1	'to taint the flavour'
			金风	*jin*T1 *feng*T1	'autumn wind'
			进风	*jin*T4 *feng*T1	'to let wind in'
			搬箱	*ban*T1 *xiang*T1	'to carry a box'
			半箱	*ban*T4 *xiang*T1	'half a box'
			生津	*sheng*T1 *jin*T1	'to promote the secretion of saliva'
			圣经	*sheng*T4 *jing*T1	'Bible'
			搬空	*ban*T1 *kong*T1	'to empty'
			半空	*ban*T4 *kong*T1	'in the air'

Non-word session				
			Pinyin	
Training Session	Targets (24 tokens)	T4.T2	*sheng xi, dao qiao, xiao kang, di qian, jian xian, shou ming, mao qi, tou chong*	
		T4.T3	*yao chun, fang jiang, jin zhun, guai jian, suan xi, bian deng, ban bian, zhuang ban*	
		T4.T4	*gua yuan, cai jin, shu xie, xian jiang, hua jian, pao di, yan chuang, fan jiao*	
	Non-targets (24 tokens)	T2.T2	*hun xing, fei qin*	
		T2.T3	*lian ping, ping fou, qiong cao*	
		T2.T4	*ya dan, qin xie, pi fen*	
		T3.T2	*jiong lu, xian qiu, mian xia*	
		T3.T3	*jian zao, fen dou*	
		T3.T4	*bing cai, fan qian, lian zheng*	
		T5.T2	*yue qian, xi qiu, qie chen*	
		T5.T3	*yao yan, ca xian, sha fan*	
		T5.T4	*za gun, te bai*	
Test Session	Targets (8 tokens)	T4.T2	*dan qiong, chuang hu, xian chao*	
		T4.T3	*qing gan, shang xing, zai qiao*	
		T4.T4	*ban dian, dong jian*	
	Non-targets (8 tokens)	T2.T2	*cong huang*	
		T2.T3	*yan chao*	
		T2.T4	*tong bao*	
		T3.T2	*qian pang*	
		T3.T3	*jie xian*	
		T3.T4	*fang kun*	
		T5.T2	*re xing*	
		T5.T3	*la lian*	
	Test tokens (16 tokens)	Sandhied T1.T1	Pinyin	Chinese characters
			*bao*T1 *qiao*T1	包敲
			*bao*T4 *qiao*T1	抱敲

			*jiang*T1 *cai*T1	将猜
			*jiang*T4 *cai*T1	降猜
			*di*T1 *guai*T1	低乖
			*di*T4 *guai*T1	地乖
			*ji*T1 *she*T1	鸡奢
			*ji*T4 *she*T1	记奢
			*zhi*T1 *tian*T1	知天
			*zhi*T4 *tian*T1	志天
			*dao*T1 *fan*T1	刀翻
			*dao*T4 *fan*T1	到翻
			*fan*T1 *cao*T1	翻操
			*fan*T4 *cao*T1	饭操
			*suan*T1 *jian*T1	酸尖
			*suan*T4 *jian*T1	算尖
			*mao*T1 *che*T1	猫车
			*mao*T4 *che*T1	帽车
			*dao*T1 *deng*T1	刀灯
			*dao*T4 *deng*T1	到灯
			*zhuang*T1 *huang*T1	装荒
			*zhuang*T4 *huang*T1	撞荒
			*cai*T1 *jin*T1	猜金
			*cai*T4 *jin*T1	菜金
			*xun*T1 *chuan*T1	熏穿
			*xun*T4 *chuan*T1	训穿
			*jiang*T1 *chuan*T1	江川
			*jiang*T4 *chuan*T1	降穿
			*guai*T1 *xiu*T1	乖修
			*guai*T4 *xiu*T1	怪修
			*pao*T1 *bian*T1	抛边
			*pao*T4 *bian*T1	泡边

Sandhi 2: T4 as the target tone

Real-word session					
			Chinese characters	Pinyin	Gloss
Training Session	Targets (24 tokens)	T4.T2	肉皮	*rou pi*	'pig skin'
			碰头	*peng tou*	'to meet'
			太平	*tai ping*	'peaceful'
			拜年	*bai nian*	'to pay a new year call'
			地坪	*di ping*	'terrace'
			豆芽	*dou ya*	'bean sprout'
			菜头	*cai tou*	'turnip'
			树皮	*shu pi*	'bark'
		T4.T3	地主	*di zhu*	'landlord'
			豆米	*dou mi*	'green soy bean'
			菜籽	*cai zi*	'vegetable seed'
			树顶	*shu ding*	'tree top'
			烫水	*tang shui*	'hot water'
			大小	*da xiao*	'size'
			阵雨	*zhen yu*	'shower'
			右手	*you shou*	'right hand'
		T4.T4	地契	*di qi*	'land contract'
			豆瓣	*dou ban*	'halves of a bean'
			菜地	*cai di*	'vegetable field'
			树洞	*shu dong*	'tree hole'
			大豆	*da dou*	'soybean'
			素菜	*su cai*	'vegetable dish'
			种树	*zhong shu*	'to plant trees'
			教授	*jiao shou*	'professor'
	Non-targets (24 tokens)	T2.T2	头皮	*tou pi*	'scalp'
			平行	*pi xing*	'parrallel'
		T2.T3	洪水	*hong shui*	'flood'
			抬手	*tai shou*	'to raise hands'

			从小	*cong xiao*	'from childhood'
		T2.T4	皮蛋	*pi dan*	'preserved egg'
			头痛	*tou tong*	'headache'
			黄豆	*huang dou*	'soybean'
		T3.T2	手头	*shou tou*	'at hand'
			眼皮	*yan pi*	'eyelid'
			水平	*shui ping*	'level'
		T3.T3	口水	*kou shui*	'saliva'
			手掌	*shou zhang*	'palm'
		T3.T4	扁豆	*bian dou*	'lentil'
			小象	*xiao xiang*	'calf elephant'
			本地	*ben di*	'local'
		T5.T2	出头	*chu tou*	'to stand out'
			雪球	*xue qiu*	'snowball'
			脚皮	*jiao pi*	'foot skin'
		T5.T3	木板	*mu ban*	'wood board'
			接手	*jie shou*	'to take over'
			喝水	*he shui*	'to drink water'
		T5.T4	赤豆	*chi dou*	'red bean'
			白菜	*bai cai*	'Chinese cabbage'
Test Session	Targets (8 tokens)	T4.T2	菜油	*cai you*	'rape oil'
			过年	*guo nian*	'to have the Spring Festival'
			太平	*tai ping*	'peaceful'
		T4.T3	放水	*fang shui*	'to drain off water'
			现场	*xian chang*	'on the spot'
			暴雨	*bao yu*	'rainstorm'
		T4.T4	做事	*zuo shi*	'to do things'
			庆祝	*qing zhu*	'to celebrate'
	Non-targets (8 tokens)	T2.T2	明年	*ming nian*	'next year'
		T2.T3	平顶	*ping ding*	'flat top'
		T2.T4	杨树	*yang shu*	'aspen tree'
		T3.T2	小河	*xiao he*	'small river'

		T3.T3	水饺	*shui jiao*	'boiled dumpling'
		T3.T4	手背	*shou bei*	'back of the hand'
		T5.T2	学年	*xue nian*	'academic year'
		T5.T3	发小	*fa xiao*	'childhood friend'
	Test tokens (16 tokens)	Sandhied T4.T5	献血	*xian*T4 *xue*T5	'to donate blood'
			先学	*xian*T1 *xue*T5	'to learn first'
			信页	*xin*T4 *ye*T5	'letter'
			新叶	*xin*T1 *ye*T5	'new leaf'
			性急	*xing*T4 *ji*T5	'impatient'
			心急	*xin*T1 *ji*T5	'impatient'
			刀滑	*dao*T4 *hua*T5	'the knife is slippery'
			倒滑	*dao*T1 *hua*T5	'backward skating'
			半秃	*ban*T4 *tu*T5	'halve bald'
			斑秃	*ban*T1 *tu*T5	'alopecia areata'
			现吃	*xian*T4 *chi*T5	'to eat on site'
			先吃	*xian*T1 *chi*T5	'to eat first'
			弹粒	*dan*T4 *li*T5	'bullet'
			单粒	*dan*T1 *li*T5	'single seed'
			化药	*hua*T4 *yao*T5	'chemical drug'
			花药	*hua*T1 *yao*T5	'anther'
			敬业	*jing*T4 *ye*T5	'delicated'
			精液	*jing*T1 *ye*T5	'seminal fluid'
			抱脚	*bao*T4 *jiao*T5	'to hold feet'
			包脚	*bao*T1 *jiao*T5	'to wrap feet'
			扇木	*shan*T4 *mu*T5	'fan made of wood'
			杉木	*shan*T1 *mu*T5	'cedarwood'
			现热	*xian*T4 *re*T5	'to heat on site'
			先热	*xian*T1 *re*T5	'to heat first'
			再活	*zai*T4 *huo*T5	'to live again'
			栽活	*zai*T1 *huo*T5	'to grow a plant'
			半桌	*ban*T4 *zhuo*T5	'half a table'
			搬桌	*ban*T1 *zhuo*T5	'to carry a table'
			幸福	*xing*T4 *fu*T5	'happiness'

			心服	*xin*T1 *fu*T5	'to be genuinely convinced'
			透血	*tou*T4 *xue*T5	'blood-soaked'
			偷学	*tou*T1 *xue*T5	'to learn secretly'

Non-word session			
			Pinyin
Training Session	Targets (24 tokens)	T4.T2	*sheng xie, dao qiao, xiao kang, di qian, jian xian, shou ming, mao qi, tou chong*
		T4.T3	*yao chun, fang jiang, jin zhun, guai jian, suan xi, bian deng, ban bian, zhuang ban*
		T4.T4	*gua yuan, cai jin, shu xie, xian jiang, hua jian, pao di, yan chuang, fan jiao*
	Non-targets (24 tokens)	T2.T2	*hun xing, fei qin*
		T2.T3	*lian ping, ping fou, qiong cao*
		T2.T4	*ya dan, qin xie, pi fen*
		T3.T2	*jiong lu, xian qiu, mian xia*
		T3.T3	*jian zao, fen dou*
		T3.T4	*bing cai, fan qian, lian zheng*
		T5.T2	*yue qian, xi qiu, qie chen*
		T5.T3	*yao yan, ca xian, sha fan*
		T5.T4	*za gun, te bai*
Test Session	Targets (8 tokens)	T4.T2	*dan qiong, chuang hu, xian chao*
		T4.T3	*qing gan, shang xing, zai qiao*
		T4.T4	*ban dian, dong jian*
	Non-targets (8 tokens)	T2.T2	*cong huang*
		T2.T3	*yan chao*
		T2.T4	*tong bao*
		T3.T2	*qian pang*
		T3.T3	*jie xian*
		T3.T4	*fang kun*
		T5.T2	*re xing*

		T5.T3	*la lian*	
	Test tokens (16 tokens)	Sandhied T4.T5	Pinyin	Chinese characters
			*ben*T4 *jie*T5	笨节
			*ben*T1 *jie*T5	奔节
			*fan*T4 *ku*T5	饭哭
			*fan*T1 *ku*T5	翻哭
			*jiao*T4 *xia*T5	叫瞎
			*jiao*T1 *xia*T5	交瞎
			*guai*T4 *shua*T5	怪刷
			*guai*T1 *shua*T5	乖刷
			*tang*T4 *cha*T5	烫插
			*tang*T1 *cha*T5	汤插
			*yao*T4 *zu*T5	要足
			*yao*T1 *zu*T5	腰足
			*fang*T4 *da*T5	放答
			*fang*T1 *da*T5	方达
			*pao*T4 *ku*T5	泡哭
			*pao*T1 *ku*T5	抛哭
			*sheng*T4 *luo*T5	盛落
			*sheng*T1 *luo*T5	生落
			*xiao*T4 *jia*T5	笑夹
			*xiao*T1 *jia*T5	消夹
			*mao*T4 *zhuo*T5	冒桌
			*mao*T1 *zhuo*T5	猫桌
			*tou*T4 *mo*T5	透末
			*tou*T1 *mo*T5	偷末
			*guai*T4 *chu*T5	怪出
			*guai*T1 *chu*T5	乖出
			*jian*T4 *gua*T5	见刮
			*jian*T1 *gua*T5	尖刮
			*yao*T4 *qia*T5	要掐
			*yao*T1 *qia*T5	腰掐
			*bian*T4 *te*T5	变特

			*bian*T1 *te*T5	边特

Sandhi 2: T1 as the target tone

Real-word session					
			Chinese characters	Pinyin	Gloss
Training Session	Targets (24 tokens)	T1.T2	花钱	*hua qian*	‘to spend money’
			天凉	*tian liang*	‘cold weather’
			山头	*shan tou*	‘mountain top’
			心情	*xin qing*	‘mood’
			真皮	*zhen pi*	‘genuine lether’
			先头	*xian tou*	‘previously’
			天平	*tian ping*	‘scale’
			新年	*xin nian*	‘new year’
		T1.T3	花粉	*hua fen*	‘pollen’
			天冷	*tian leng*	‘cold weather’
			山谷	*shan gu*	‘valley’
			心口	*xin kou*	‘heart’
			清水	*qing shui*	‘clear water’
			春晓	*chun xiao*	‘early spring’
			春雨	*chun yu*	‘spring rain’
			亲手	*qin shou*	‘by hand’
		T1.T4	开放	*kai fang*	‘to open’
			天亮	*tian liang*	‘dawn’
			山芋	*shan yu*	‘sweet potato’
			心算	*xin suan*	‘mental arithmetic’
			荒地	*huang di*	‘waste land’
			豌豆	*wan dou*	‘pea’
			菠菜	*bo cai*	‘spinach’
			栽树	*zai shu*	‘to plant trees’
	Non-	T2.T2	头皮	*tou pi*	‘scalp’

	targets (24 tokens)		平行	*pi xing*	'parrallel'
		T2.T3	洪水	*hong shui*	'flood'
			抬手	*tai shou*	'to raise hands'
			从小	*cong xiao*	'from childhood'
		T2.T4	皮蛋	*pi dan*	'preserved egg'
			头痛	*tou tong*	'headache'
			黄豆	*huang dou*	'soybean'
		T3.T2	手头	*shou tou*	'at hand'
			眼皮	*yan pi*	'eyelid'
			水平	*shui ping*	'level'
		T3.T3	口水	*kou shui*	'saliva'
			手掌	*shou zhang*	'palm'
		T3.T4	扁豆	*bian dou*	'lentil'
			小象	*xiao xiang*	'calf elephant'
			本地	*ben di*	'local'
		T5.T2	出头	*chu tou*	'to stand out'
			雪球	*xue qiu*	'snowball'
			脚皮	*jiao pi*	'foot skin'
		T5.T3	木板	*mu ban*	'wood board'
			接手	*jie shou*	'to take over'
			喝水	*he shui*	'to drink water'
		T5.T4	赤豆	*chi dou*	'red bean'
			白菜	*bai cai*	'Chinese cabbage'
Test Session	Targets (8 tokens)	T1.T2	周年	*zhou nian*	'anniversary'
			加钱	*jia qian*	'to add money'
			书房	*shu fang*	'study'
		T1.T3	车顶	*che ding*	'car roof'
			家长	*jia zhang*	'parents'
			冲水	*chong shui*	'to flush'
		T1.T4	青菜	*qing cai*	'green vegetables'
			松树	*song shu*	'pine tree'
	Non-targets	T2.T2	明年	*ming nian*	'next year'
		T2.T3	平顶	*ping ding*	'flat top'

	(8 tokens)	T2.T4	杨树	*yang shu*	'aspen tree'
		T3.T2	小河	*xiao he*	'small river'
		T3.T3	水饺	*shui jiao*	'boiled dumpling'
		T3.T4	手背	*shou bei*	'back of the hand'
		T5.T2	学年	*xue nian*	'academic year'
		T5.T3	发小	*fa xiao*	'childhood friend'
	Test tokens (16 tokens)	Sandhied T4.T5	献血	*xian*T4 *xue*T5	'to donate blood'
			先学	*xian*T1 *xue*T5	'to learn first'
			信页	*xin*T4 *ye*T5	'letter'
			新叶	*xin*T1 *ye*T5	'new leaf'
			性急	*xing*T4 *ji*T5	'impatient'
			心急	*xin*T1 *ji*T5	'impatient'
			刀滑	*dao*T4 *hua*T5	'the knife is slippery'
			倒滑	*dao*T1 *hua*T5	'backward skating'
			半秃	*ban*T4 *tu*T5	'halve bald'
			斑秃	*ban*T1 *tu*T5	'alopecia areata'
			现吃	*xian*T4 *chi*T5	'to eat on site'
			先吃	*xian*T1 *chi*T5	'to eat first'
			弹粒	*dan*T4 *li*T5	'bullet'
			单粒	*dan*T1 *li*T5	'single seed'
			化药	*hua*T4 *yao*T5	'chemical drug'
			花药	*hua*T1 *yao*T5	'anther'
			敬业	*jing*T4 *ye*T5	'delicated'
			精液	*jing*T1 *ye*T5	'seminal fluid'
			抱脚	*bao*T4 *jiao*T5	'to hold feet'
			包脚	*bao*T1 *jiao*T5	'to wrap feet'
			扇木	*shan*T4 *mu*T5	'fan made of wood'
			杉木	*shan*T1 *mu*T5	'cedarwood'
			现热	*xian*T4 *re*T5	'to heat on site'
			先热	*xian*T1 *re*T5	'to heat first'
			再活	*zai*T4 *huo*T5	'to live again'
			栽活	*zai*T1 *huo*T5	'to grow a plant'
			半桌	*ban*T4 *zhuo*T5	'half a table'

			搬桌	*ban*T1 *zhuo*T5	‘to carry a table’
			幸福	*xing*T4 *fu*T5	‘happiness’
			心服	*xin*T1 *fu*T5	‘to be genuinely convinced’
			透血	*tou*T4 *xue*T5	‘blood-soaked’
			偷学	*tou*T1 *xue*T5	‘to learn secretly’

Non-word session			
			Pinyin
Training Session	Targets (24 tokens)	T1.T2	*dong qian, shang ya, guo que, hua qiong, shi qie, chuan qin, quan ting, yin he*
		T1.T3	*shou xiang, cai xian, bao yan, xian xiao, pao lian, ban huo, ji bian, chun jiang*
		T1.T4	*yan jun, xiao tang, zeng lian, hui jiang, qian dian, fen xiang, jian fei, wan bian*
	Non-targets (24 tokens)	T2.T2	*hun xing, fei qin*
		T2.T3	*lian ping, ping fou, qiong cao*
		T2.T4	*ya dan, qin xie, pi fen*
		T3.T2	*jiong lu, xian qiu, mian xia*
		T3.T3	*jian zao, fen dou*
		T3.T4	*bing cai, fan qian, lian zheng*
		T5.T2	*yue qian, xi qiu, qie chen*
		T5.T3	*yao yan, ca xian, sha fan*
		T5.T4	*za gun, te bai*
Test Session	Targets (8 tokens)	T1.T2	*jiao cao, tou fang, fen po*
		T1.T3	*xin fen, xian bao, chao bing*
		T1.T4	*pin qian, gou hui*
	Non-targets (8 tokens)	T2.T2	*cong huang*
		T2.T3	*yan chao*
		T2.T4	*tong bao*
		T3.T2	*qian pang*

		T3.T3	*jie xian*	
		T3.T4	*fang kun*	
		T5.T2	*re xing*	
		T5.T3	*la lian*	
	Test tokens (16 tokens)	Sandhied T4.T5	Pinyin	Chinese characters
			*ben*T4 *jie*T5	笨节
			*ben*T1 *jie*T5	奔节
			*fan*T4 *ku*T5	饭哭
			*fan*T1 *ku*T5	翻哭
			*jiao*T4 *xia*T5	叫瞎
			*jiao*T1 *xia*T5	交瞎
			*guai*T4 *shua*T5	怪刷
			*guai*T1 *shua*T5	乖刷
			*tang*T4 *cha*T5	烫插
			*tang*T1 *cha*T5	汤插
			*yao*T4 *zu*T5	要足
			*yao*T1 *zu*T5	腰足
			*fang*T4 *da*T5	放答
			*fang*T1 *da*T5	方达
			*pao*T4 *ku*T5	泡哭
			*pao*T1 *ku*T5	抛哭
			*sheng*T4 *luo*T5	盛落
			*sheng*T1 *luo*T5	生落
			*xiao*T4 *jia*T5	笑夹
			*xiao*T1 *jia*T5	消夹
			*mao*T4 *zhuo*T5	冒桌
			*mao*T1 *zhuo*T5	猫桌
			*tou*T4 *mo*T5	透末
			*tou*T1 *mo*T5	偷末
			*guai*T4 *chu*T5	怪出
			*guai*T1 *chu*T5	乖出
			*jian*T4 *gua*T5	见刮
			*jian*T1 *gua*T5	尖刮

			*yao*T4 *qia*T5	要掐
			*yao*T1 *qia*T5	腰掐
			*bian*T4 *te*T5	变特
			*bian*T1 *te*T5	边特

Appendix E: Instructions for the Concept Formation experiment

English instructions on the screen (to NL participants)

(adapted from Jaeger (1986)):

Before training session:

"You will be listening to a series of spoken disyllabic words in an unknown language. Some of them will have a certain melodic property in a fixed position, and others will not. After each word there will be a 3-second pause, and then the voice on the tape will tell you whether the word you just heard had the melody or not ('target/non-target'). Your job is to figure out what the melodic property is that all the 'target' words have in common and that the 'non-target' words don't have. You should click the 'target' or the 'non-target' key as quickly as possible, before the voice on the tape says the correct answer. A shrinking green bar on top of the screen will tell you how much time you have left to make the response.

In the first few trials you may simply do random clicking if you have no idea about what is 'target' and what is 'non-target.' Just do the clicking. After a few trials you may gradually get what a 'target' and 'non-target' item should be like and you will begin real selection. You will know that you have figured out the correct melodic property when your responses always match those of the voices on the tape. You will pass the training session when you perform 13 trials in a row with ≤ 2 errors. Remember that you are listening for some melodic properties that all the 'target' words have in common and that the 'non-target' words don't have."

Before test session:

"This was the end of the training session. Now the test session starts. You are going to hear 32 words, and your task is to respond in the same way as before. The target melody you are looking for is the same as you have just found. During the test you will receive no feedback. Again, make sure you respond before the shrinking green bar disappears."

Chinese instructions on the sreen (to BJ and NJ listeners):

Before training session:

“您将听到一组来自某方言的双字组词语。这些词语中，有一些在词语的固定位置有某种声调，而另一些没有这一声调。每个词语后有 3 秒的停顿，然后耳机里的提示音将告诉您刚刚听到的词语中是否有这种声调("目标/非目标")。您的任务是识别目标声调。当您听到词语后，请在提示音给出正确答案前尽快按下“目标”按钮或者“非目标”按钮。屏幕上方的绿色时间条会提示本题剩下的作答时间。

在听到前几个词语时，因为还没有找到目标声调，您可能只是随机按下按钮。在听了几个词语之后，您会慢慢熟悉目标声调。如果您的选择总是和提示音一致，这说明您正确识别了目标声调。当您在连续 13 个词语中有 11 个或以上选择正确的时候，您将通过本训练阶段。”

Before test session:

"训练阶段结束。现在开始测试阶段。您将听到 32 个词语，请您像刚才一样听到词语后按下按钮。您的目标声调仍是刚才的声调，但本阶段将不会播放正确答案的提示音。同样，请在绿色时间条消失前按下按钮。”

Appendix F: Full results from the Tukey post-hoc test in Chapter 3

Linear Hypotheses	Estimate	Standard error	z	p
NJsurface – BJsurface == 0	-1.81	0.28	-6.52	< .001 ***
NLsurface – BJsurface == 0	-2.70	0.28	-9.81	< .001 ***
BJunderlying – BJsurface == 0	**-5.31**	**0.30**	**-17.77**	**< .001** ***
NJunderlying – BJsurface == 0	-1.82	0.28	-6.57	< .001 ***
NLunderlying – BJsurface == 0	-2.59	0.28	-9.42	< .001 ***
NLsurface – NJsurface == 0	-0.89	0.25	-3.49	< .01 **
BJunderlying – NJsurface == 0	-3.50	0.28	-12.57	< .001 ***
NJunderlying – NJsurface == 0	**-0.01**	**0.26**	**-0.05**	**1.000**
NLunderlying – NJsurface == 0	-0.78	0.26	-3.07	< .05 *
BJunderlying – NLsurface == 0	-2.61	0.27	-9.55	< .001 ***
NJunderlying – NLsurface == 0	0.88	0.25	3.46	< .01 **
NLunderlying – NLsurface == 0	**0.11**	**0.25**	**0.43**	**0.998**
NJunderlying – BJunderlying == 0	3.48	0.28	12.58	< .001 ***

NLunderlying – BJunderlying == 0	2.71	0.27	9.93	< .001 ***
NLunderlying – NJunderlying == 0	-0.77	0.25	-3.03	< .05 *

Appendix G: Instructions for the Word Detection experiment

The instructions are displayed on the computer screen:

> "In this part you are going to hear words in pairs. The first word is always a monosyllabic word, and the second word is a disyllabic word. Your task is to indicate as quickly as possible whether you think in each pair the first word is repeated in the second word. Only press the *YES* button when the first word is repeated in the second word with the correct melody!"

Nederlandse samenvatting

Het voornaamste doel van dit onderzoek is om de taalspecifieke en algemene kennis te onderzoeken die luisteraars helpt bij het reconstrueren van onderliggende representaties uit fonetische representaties, hier aangeduid als een 'mapping'. De algemene onderzoeksvragen zijn: 1) Is de mapping uniek afhankelijk van taalspecifieke kennis, of wordt deze ook gefaciliteerd door taalonafhankelijke kennis van articulatie? 2) Verschilt de moeilijkheid van mapping voor assimilatie- en dissimilatieprocessen, in het bijzonder voor naïeve niet-moedertaalluisteraars? 3) Wordt de mapping beïnvloed door de mate waarin assimilatie- en dissimilatieprocessen compleet of gradueel zijn, voor naïve niet-moedertaalluisteraars?

De taalspecifieke versus algemene kennis die nodig is voor de mapping van onderliggende representaties uit fonetische representaties wordt hier onderzocht via assimilatie- en dissimilatieprocessen. Assimilatie-processen vinden hun oorsprong in gemak van articulatie. Zowel de 'Motor Theory' als de 'Direct Realist Theory' gaan ervan uit dat akoestische spraakverschijnselen worden waargenomen als zijnde veroorzaakt door articulatorische 'gestures'/gebaren. De impliciete kennis over gestures, gegeneraliseerd uit ervaring van de spreker met articulatie van de moedertaal of uit algemene articulatorische kennis, is niet beperkt tot een bepaalde taal, maar taalonafhankelijk. Hieruit volgt de hypothese dat mapping van onderliggende uit fonetische representaties voor assimilatieprocessen waarschijnlijk eveneens taalonafhankelijk is (de 'taalonafhankelijke mapping-hypothese voor assimilatieprocessen'). Dissimilatieprocessen daarentegen worden meestal verklaard door de 'hypercorrectie'-theorie of, minder vaak, door de 'motor planning'-theorie. De 'hypercorrectie'-theorie impliceert dat dissimilatieprocessen intrinsiek moeten refereren aan onderliggende representaties, waarover moedertaalluisteraars beschikken, maar die niet beschikbaar zijn voor naïeve niet-moedertaalluisteraars. Vandaar dat volgens dit standpunt de mapping van onderliggende uit fonetische representaties voor dissimilatieprocessen alleen toegankelijk is voor moedertaalluisteraars en ontoegankelijk voor naïeve niet-moedertaalluisteraars (de 'taalspecifieke mapping-hypothese voor dissimilatieprocessen'). Echter, het 'motor planning'-standpunt doet geen eenduidige en consistente voorspellingen over de taalspecificiteit/taalonafhankelijkheid van mapping voor dissimilatieprocessen. Dit is waarom in deze dissertatie wordt aangenomen dat de taalspecifieke mapping-hypothese voor dissimilatieprocessen alleen door het 'hypercorrectie'-standpunt eenduidig wordt voorspeld.

Als aanvulling op de taalonafhankelijke mapping-hypothese voor assimilatieprocessen die gebaseerd is op articulatorische gestures, wordt voorgesteld dat de mapping voor assimilatieprocessen wellicht alleen beschikbaar is voor niet-moedertaalluisteraars in het geval dat de onderliggende representatie in zekere mate akoestisch afleidbaar is uit de fonetische representatie (de 'graduele taalonafhankelijke mapping-hypothese voor assimilatieprocessen').

Deze dissertatie richt zich op toon- en toonsandhi-verschijnselen om de bovenstaande hypothesen over de mapping van onderliggende uit fonetische representaties te onderzoeken. Omdat luisteraars van niet-toontalen maximaal verschillen van moedertaalluisteraars van toontalen wat betreft hun taalspecifieke articulatorische kennis van toon, vormen zij een ideale groep van luisteraars om het effect te testen van taalonafhankelijke articulatorische kennis in mapping.

Hoofdstuk 2 is een productie-experiment van het Nanjing Mandarijn. Dit onderzoek was gericht op het akoestisch onderzoeken van de productie van twee paren van vergelijkbare assimilatie- en dissimilatie- toonsandhi-processen in het Nanjing Mandarijn zoals beschreven in eerdere literatuur, om deze beschrijvingen te verifiëren en te beslissen of de sandhi-verschijnselen gebruikt kunnen worden als basis voor de stimuli voor de mapping-studie in Hoofdstuk 3. Het onderzoek vergeleek de toonsandhi-patronen geproduceerd door moedertaalsprekers van het Nanjing Mandarijn met de fonetische beschrijvingen in eerdere studies. Een paar sandhi-processen (een assimilatie en een dissimilatie) werd geschikt bevonden om gebruikt te worden als basis voor de stimuli voor de mapping-studie in Hoofdstuk 3.

Als vrijstaande studie beoogt Hoofdstuk 2 ten eerste het testen van de assumptie dat assimilatieprocessen uit articulatiegemak voortkomen, welke voorspelt dat dergelijke processen gradueel kunnen zijn; en ten tweede, vergelijken van de 'hypercorrectie'-theorie versus de 'motor planning'-theorie over de motivatie voor dissimilatieprocessen, waarvan alleen de eerste eenduidig het categoriale karakter van dissimilatie-processen voorspelt. Het bleek dat twee geselecteerde assimilatorische toonsandhi-processen verschillen qua categoriaal/gradueel karakter, wat in overeenstemming is met de voorspelling vanuit de assumptie van articulatiegemak. Verder werd geobserveerd dat ook de twee dissimilatorische toonsandhi-processen verschillen qua categoriaal/gradueel karakter, wat niet in overeenstemming is met de voorspelling van de 'hypercorrectie' theorie.

Hoofdstuk 3 is een mapping-studie die de twee algemene onderzoeksvragen 1) en 2) behandelt, gebruik makend van het Nanjing Mandarijn als een tonale doeltaal, en met Concept Formatie als experimenteel paradigma. De studie omvatte een groep moedertaalsprekers van het Nanjing Mandarijn; een naïeve niet-toontaal groep van Nederlandse luisteraars; en een Beijing Mandarijn groep, die dient als niet-moedertaal toontaalgroep, dus als controlegroep. De resultaten laten zien dat alleen de Nanjing moedertaalluisteraars succesvol waren in de mapping van fonetische representaties naar onderliggende representaties, in zowel de assimilatie- als de dissimilatie-sandhiprocessen van het Nanjing Mandarijn. Hun fonologische kennis en/of lexicale kennis van de moedertaal stelde hen in staat om de contextuele fonetische neutralisatie van tonen ongedaan te maken, en aldus de mapping van de fonetische sandhitoon naar de onderliggende toon te volbrengen, zoals voorspeld. De naïeve Nederlandse luisteraars slaagden er niet in gebruik te maken van de assimilatiecontext om een articulatiegebaseerde link te leggen tussen fonetische en onderliggende tonen, hetgeen de voorspelling tegenspreekt. In plaats daarvan presteerden de luisteraars rond kansniveau op zowel mapping van fonetische naar onderliggende tonen, als op de onderlinge correspondentie tussen fonetische tonen. Deze luisteraars vormden waarschijnlijk middels een korte intensieve toontrainingstaak tijdelijke onstabiele categorieën/quasi-categorieën voor Nanjing Mandarijn tonen, maar ze ondervonden problemen in het vasthouden van deze categorieën tijdens de erop volgende testsessie, waarin ze niet in staat waren ze te gebruiken in de mapping-taak. Deze interpretatie werd bevestigd door een vergelijking met de controlegroep van luisteraars met Beijing Mandarijn-achtergrond.

De resultaten van deze studie bevestigen de rol van taalspecifieke kennis in het uitvoeren van mappings van fonetische naar onderliggende vormen door moedertaalluisteraars. Echter, de resultaten impliceren niet dat taalonafhankelijke articulatorische kennis helemaal geen rol heeft in het mogelijk maken van het uitvoeren van mapping van fonetische naar onderliggende vormen door naïeve moedertaalluisteraars. De taak die werd gebruikt in deze studie was wellicht niet cognitief uitdagend genoeg voor de Nederlandse luisteraars, zowel toen zij getraind werden om een tooncategorie te leren in een korte tijd (ongeveer 20 minuten) als toen ze vervolgens gevraagd werd om deze tijdelijk aangeleerde tooncategorie te gebruiken in een mapping-taak. Om deze reden vraagt een relatief kortdurende experimentele studie om een ander type experimenteel paradigma voor het testen van Nederlandse luisteraars.

Hoofdstuk 4 is een tweede studie naar de mapping van fonetische tonen naar onderliggende tonen door naïeve niet-moedertaalluisteraars. Het onderzocht net als Hoofdstuk 3 de algemene onderzoeksvraag 2), waarbij een cognitief minder uitdagende taak gebruikt werd, namelijk 'Word Detection'. Verder werd de algemene onderzoeksvraag 3) onderzocht. Experiment 1 in deze studie vergeleek mapping van fonetische naar onderliggende tonen voor categoriale assimilatorische en dissimilatorische toonsandhi-processen en verschafte geen bewijs waaruit blijkt dat de Nederlandse participanten succesvoller zijn in mapping in de assimilatieconditie vergeleken met de dissimilatieconditie. Experiment 2 focuste meer op graduele toonsandhi; er bleek dat luisteraars een betere mapping vertoonden in de assimilatieconditie op basis van twee soorten data: een verschil in de reactietijden en een trend van verschil in de detectiewaarden. Deze resultaten kunnen geïnterpreteerd worden als enig bewijs voor de graduele taalonafhankelijke mapping-hypothese voor assimilatieprocessen.

De bevinding dat naïeve niet-moedertaal luisteraars alleen beter waren in mappings van fonetische naar onderliggende tonen voor graduele toonsandhi-processen suggereert niet noodzakelijk dat categoriale assimilatieprocessen naïeve niet-moedertaalluisteraars zouden hinderen in het uitvoeren van mappings van fonetische tonen naar onderliggende tonen. In plaats daarvan is het waarschijnlijk dat de Nederlandse luisteraars in Experiment 1 geen mapping-voordeel voor categoriale assimilatieprocessen vertoonden omdat ze werden afgeleid door versterkte aanwezigheid van cues voor toonsegmentatie binnen dezelfde taak (in de dissimilatieconditie). In Experiment 1 hadden de luisteraars wellicht een perceptief voordeel van categoriale dissimilatieprocessen doordat de segmentatie-cue intrinsiek aan deze processen ervoor zorgde dat ze de fonetische toon beter konden localiseren in de context, waardoor de fonetische toon een transparantere correspondentie had met de onderliggende toon; terwijl ze minder gefaciliteerd leken in het segmenteren van de fonetische toon uit de context in de categoriale assimilatieconditie, waarin deze segmentatie-cues ontbraken. Het dissimilatieproces dat gebruikt werd in Experiment 1 bevatte bovendien een herhaling in de fonetische toonreeks, wat ook localisatie kan faciliteren. Experiment 3 onderzocht of de verschijnselen geobserveerd in Experiment 2 veroorzaakt werden door een abrupte F0-incongruentie in de toonreeks (die intrinsiek is aan categoriale dissimilatieprocessen) of door de herhaalde toonreeks; beide factoren kunnen het geobserveerde mapping-voordeel in de dissimilatieconditie voor Nederlandse deelnemers verklaren, maar de twee effecten bleven uiteindelijk niet van elkaar onderscheidbaar; echter, de abrupte F0-incongruentie was waarschijnlijk de hoofdoorzaak.

Concluderend, deze dissertatie heeft bewijs geleverd voor taalspecificiteit (in de zin van afhankelijkheid van luisteraars van hun moedertaalfonologie) alsmede voor taalonafhankelijkheid (in de zin van gebruik door luisteraars van algemene kennis van articulatorische gestures) van de mapping van fonetische naar onderliggende tonen voor assimilatorische en dissimilatorische toonsandhi-processen. Bovendien wijzen de resultaten erop dat taalonafhankelijke segmentatiefactoren een rol spelen in mappings van categoriale dissimilatieprocessen (althans in vergelijking met categoriale assimilatieprocessen).

Curriculum Vitae

Xin Li was born on February 3, 1989 in Anhui, China. She obtained her BA in English in 2010 from China Pharmaceutical University, Nanjing China, and her MA in Linguistics and Applied Linguistics in 2013 from Tongji Univesity, Shanghai, China. She started her PhD project at Utrecht University in December 2013. This dissertation is the result of her PhD research.