Universiteit Leiden
The Netherlands

# Production of Tone 3 Sandhi in Reduplication and Compounds

Knuppenburg, Amanda

**Citation**
Knuppenburg, A. (2024). *Production of Tone 3 Sandhi in Reduplication and Compounds*.

Version: Not Applicable (or Unknown)
License: License to inclusion and publication of a Bachelor or Master Thesis, 2023
Downloaded from: https://hdl.handle.net/1887/3775085

**Note:** To cite this publication please use the final published version (if applicable).

# Production of Tone 3 Sandhi in Reduplication and Compounds

Amanda Knuppenburg (s3892085)

MA Linguistics, Leiden University

MA Thesis

Supervisor Dr. T.J. Laméris

July 1, 2024

# Table of Contents

# Abstract

Tone three sandhi is a tonal alteration that occurs when two T3 syllables are in an immediate environment of each other /T3 + T3/ but is pronounced with the first syllable as a T2 on the surface [T2 + T3] (Duanmu, 2007). From a previous processing experiment (Gao et al., 2021), differences in how this sandhi is perceived were found between two differing morphological environments: reduplicated and compound verbs. This thesis attempts to understand the underlying and surface forms by analyzing the production of tone three sandhi in these two morphological constructions. It further attempts to understand the complex relationship between perception and production through the avenue of converting a previous speech perception experiment (Gao et al., 2021) into a production one. It predicts that the initial morpheme in sandhi compound verbs will be more like a standard T2 than in sandhi reduplicated verbs based on the differences found in previous research. Using similar stimuli to Gao et al. (2021), participants were asked to produce sentences containing the sandhi stimuli to later analyze the pitch contours through contour clustering software (Kaland, 2021). The results showed limited differences between the compound and reduplicated verbs. Concluding that, given the data, the difference in constructing morphology of compound and reduplicated verbs is not present in the produced tones. Furthermore, speech perceived differences between these two morphological constructions and sandhi does not have any evidence of produced differences identifying a barrier between perception and production.

# 1 Background

A typically understudied phenomenon in Chinese is tone three sandhi. While there are a few perception experiments about this tonal alteration (Li et al., 2016; W. Wang and Li, 1967; Gao et al., 2021; etc.), there are relatively few production experiments.

From a previous processing experiment (Gao et al., 2021), differences in a morphological environment — reduplicated and compound verbs — were found to affect how easily a native speaker perceives tone sandhi. This is claimed to be due to underlying and surface shifts in sandhi in combination with the morphological structures. With this background, this thesis will shift a perception experiment into a production one to further test these different morphological structures and if the perception differences affect production.

## 1.1 Chinese Tones

Chinese is a language that can be divided up into many dialects. Despite the term 'dialect' used for the varieties of languages spoken in China, natives to each area find it difficult to understand other dialects due to their phonology. Still, they all use the same written language and grammar (Duanmu, 2007) which has all dialects considered under the same branch of language: Chinese.

While the phonology across dialects can be vastly different, there have been several language reform movements to shift people to use Standard Chinese — also called Putonghua — to ease communication across different dialects. Standard Chinese is based on the Mandarin family dialects. Regional dialect use is still present, but most Chinese people will understand and speak Standard Chinese (Duanmu, 2007). However, it is not everyone's native language. Even people who think they natively speak Standard Chinese can still have an 'accent' when speaking as, "people spend only as much effort learning [Standard Chinese] as will make them understood, and do not bother with the accent they still have," (Duanmu, 2007, pg 4). Therefore, people not from the north of China or from a predominantly Mandarin-speaking region tend to lack perfect pronunciation of Standard Chinese.

In Standard Chinese there are four tones: (T1) high, (T2) rising, (T3) low, and (T4) falling. A visualization of the pitch contours is below in Figure 1:

**Figure 1**

*Tone Contours of Standard Chinese (Xu, 1997)*

*Contextual tonal variations*

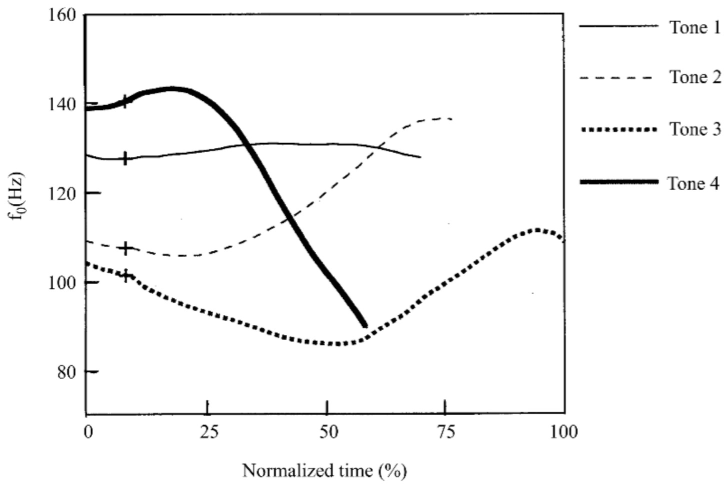



**Figure 2.** Mean $f_0$ contours (averaged over speakers and tokens; $n = 48$) of four Mandarin tones in the monosyllable /ma/ produced in isolation. The time is normalized, with all tones plotted with their average duration proportional to the average duration of Tone 3.

(Xu, 1997)

Figure 1 is a normalized graph of the four Mandarin tones on the monosyllable /ma/ produced in isolation. Across the x-axis is the normalized time with some tones being shorter and longer. The y-axis is pitch (or f0), which shows the fluctuation of each tone over the normalized time.

Tone 1 (T1) is a high tone that is maintained and constant across the vowel. Figure 1 shows this with a thin solid black line that stays constant over the duration of the vowel. Tone 2 (T2) on the other hand, is a more fluctuating pitch shown by a dashed black line that starts lower then goes high over the duration of the vowel. This is known as a rising tone.

Tone 3 (T3) is visualized by the dotted black line that is the longest pitch contour. This is typically known as the ‘low’ tone in most research done on Chinese (Gao et al., 2021), reasoned as it does have the lowest tone contour. However, the pitch is not constantly low — like how T1 is constantly high — Tone 3 fluctuates in more of a dipping motion: starting high, dipping low and then rising back up.

Meanwhile, Tone 4 (T4) is the shortest tone in Chinese, visualized in Figure 1 by the thickest black line. This is the opposite of Tone 2 where Tone 4 starts high and ends low.

For the most part, all morphemes in Chinese have a tone on their vowels that give it semantic meaning. Consequently, the same pair of consonants and vowels with different tones will sound similar to the untrained ear but have entirely different meanings. An example of this is below:

**Table 1**
*Mandarin Tones*

| | | | | |
|---|---|---|---|---|
| Pinyin | mā | má | mǎ | mà |
| Character | 妈 | 痲 | 马 | 骂 |
| Tone | T1 | T2 | T3 | T4 |
| Gloss | “mother” | “hemp” | “horse” | “scold” |

The only difference across Table 1 is the tone. They all are /ma/ but with differing tones that change the meaning from ‘mother’ (T1) to ‘horse’ (T3) and so on.

These four tones are utilized in most words in Chinese however, there is a neutral tone called Tone 0 (T0) that is used for more particle words in Chinese rather than content words (Shen, 2022). For instance, Tone 0 is used the structural particle words like /的/ ‘de’. In Chinese, /的/ ‘de’ is often used to indicate possession or indicating an aspect of a noun. For instance, in the phrase “漂亮的花” (piàoliang de huā) which translates to ‘beautiful flower’. This has the adjective /漂亮/ ‘piàoliang’ meaning ‘beautiful’ and the noun /花/ ‘huā’ meaning ‘flower’ connected with /的/ ‘de’ which attaches the two words showing that the flower has the aspect of being beautiful. Now, structural words are not the only words without tone, morphological transformations — such as reduplication — may result in a usual content word losing its tone and becoming T0 (Duanmu, 2007).

There are instances and environments in Chinese that change tones from how they are thought about in speakers’ minds, to how they are actually pronounced. Speaker’s may think they are saying one tone (underlying), but in reality, they are saying another (surface). This alteration happens with environments pertaining to Tone 3 and called Tone Three Sandhi and with Tone 0 with certain morphological categories — reduplication.

## 1.2 Underlying and Surface Forms

The way people think about language and how they pronounce it, do not always match up with how people actually pronounce words. This is known as a difference between underlying forms — how a person thinks something is pronounced — and surface forms — how it is actually pronounced. The underlying form tends to be more of an abstract representation that the surface form can be derived from (Lahirir & Marslen-Wilson, 1991). The surface forms can be derived through a series of transformations that the underlying form triggers.

In Mandarin, one of these transformations is called tone three sandhi.

### *1.2a Tone Three Sandhi*

Tone three sandhi has been observed as early as the sixteenth century (Mei, 1977) and has been studied by many researchers since. Theoretical work takes the forefront of research in tone three sandhi (Duanmu, 2007, Cheng, 1973, Shen, 1994, Shih, 1986, 1997, Chen, 2000, etc.) with limited experimentation done to provide quantitative data about the intricacies of this alternation in the minds of natives.

Tone three sandhi is a tonal alteration that occurs when two T3 syllables are in an immediate environment of each other /T3 + T3/, when this is produced by a native speaker, the first syllable changes to a T2 being realized on the surface as [T2 + T3] (Duanmu, 2007). This happens across words, parts of speech, phrases, and many more environments in Standard Chinese, but for the purposes of this thesis, the focus will be on compound and reduplicated verbs to be discussed shortly.

Tone sandhi does not just affect Standard Chinese, there are other dialects and even other tonal languages that experience this type of alternation like Vietnamese and Thai — but in their own variation of tones (Xu, 1997). Furthermore, within the many dialects of Chinese itself there are dialects with more tones than others that experience sandhi differently than Mandarin (Duanmu, 2005). Furthermore, there are other tone alternations in Standard Chinese that have been studied in combination with tone three sandhi to understand how productive different transformations are in relation to each other (Zhang & Lai, 2010).

Zhang and Lai (2010) performed an experiment with the ‘wug’ test in Mandarin Chinese. A ‘wug’ test is essentially testing native speakers’ pronunciation of fake words (or ‘wug’s) to understand how productive phonological rules known on real words are when pronouncing fake ones. In Zhang and Lai (2010)’s case, they had participants listen and repeat a combination of real and fake words to analyze the different tonal contours for a variety of disyllabic words. They were testing two patterns of tonal alternation (tone sandhi) that differed on what they claimed was motivation — tone three sandhi and half-third sandhi.

In combination of tone three sandhi, they also were looking at another tone sandhi called ‘half-third sandhi’ where tone three (T3) becomes a half tone three. As discussed earlier (recall Table 1), tone three (T3) has the pitch contour of a dipping motion like high-low-high, and what Zhang and Lai (2010) claimed happens when T3 is cut in half is the later ‘high’ — or rising motion — disappears. Therefore, the T3 pitch contour becomes high-low, similar to a tone four (T4) but is overall much lower than a true T4 (as T4’s initial f0 is higher than T3’s initial f0) and therefore in this thesis the half-third sandhi tonal alteration will be defined as a normal T3 becoming a weak T4 when followed by any other tone except tone three.

To reiterate the two tonal transformations discussed: tone three sandhi is when the underlying environment /T3 T3/ triggers a tonal alternation and becomes [T2 T3], meanwhile half-third sandhi is when the underlying environment /T3 TX/ (X being any tone besides T3) will cut the initial T3 contour so that it is a weak T4 becoming [T4 TX].

In Zhang and Lai (2010)’s production study using the ‘wug’ test they claimed that tone three sandhi applies consistently to ‘wug’ words, “but the application is incomplete, in that the sandhi tone bears more resemblance to the base tone than the sandhi tone in real words,” (Zhang & Lai, 2010, pg 186). This means that in the production of real words, the tone three sandhi appeared as expected on the surface as [T2 T3]. Meanwhile in 'wug' words, the initial tone resembled more of the base tone, a /T3/, which Zhang and Lai (2010) labeled as incomplete application.

However, it is interesting that the pronunciation of sandhi differs in this way, where the productivity is incomplete in fake words. Perhaps more production studies on sandhi can reveal if there are other differences between expected tone sandhi production and how it is actually pronounced. This thesis will attempt to find as such in two different morphological structures in Standard Chinese: Reduplicated and Compound verbs.

### *1.2b Morphology*

To fully understand these two structures of verbs, one must become familiar with the basic morphology of how Chinese words and syllables are organized. For the most part, Chinese words can be split into two categories: monosyllabic and polysyllabic words (Yu, 2003).

Words can be divided into small bits of meaning. These are called morphemes — the smallest sound that holds meaning in a given language. In Chinese, monosyllabic is just one syllable that is a morpheme. For instance, [人] 'rén' means 'person' and can stand alone in a sentence. Polysyllabic words can be composed of two or more morphemes like [十字路口] 'shízìlù kǒu' which translates to 'intersection' and contains many different morphemes — such as [十字] 'shízì' which means 'cross' and [路] 'lù' which means 'road' — or polysyllabic words can also be one morpheme but multiple syllables such as the word for 'marathon' in Chinese is [马拉松] 'mǎlāsōng' (Yu, 2003). These later categories of Chinese polysyllabic words are often borrowed from other languages therefore the meanings behind each individual character do not necessarily relate to the meaning but is phonologically similar. For instance, in [马拉松] the first character [马] 'mǎ' means 'horse', the second [拉] 'lā' means 'pull' and [松] 'sōng' means 'loose' but together the phonology roughly resembles the English word 'marathon' with the pinyin being 'mǎlāsōng'.

Furthermore, Chinese makes use of the morphological function of reduplication across all parts of speech. Reduplication is when all or part of a lexical item is repeated. For instance, the verb [看] 'kàn' which means 'to watch', can be reduplicated into [看看] 'kàn kan' now meaning 'to watch for a little bit'. When the entire original sound 'kàn' is repeated to become 'kàn kan', this is called full reduplication. There is also partial reduplication such as [评理] 'píng lǐ' meaning 'to judge' becoming [评评理] as 'píng píng lǐ' now meaning 'to try and judge' (Chen et al., 1992). Here only 'píng' is being reduplicated and not 'lǐ', making it partially reduplicated.

With this understanding of some features of Chinese morphology, there are even some more transformations that occur within them. Not only can a word trigger one transformation — such as the sandhi context of two T3's next to each other — it can oftentimes trigger multiple as sandhi is not the only transformation to occur in Chinese. Reduplication also has some tonal transformations to be discussed shortly that are separate from sandhi but can be related if the environment's overlap. For the most part — and in the context of this thesis and its verbs — this transformation is a tone dropping of the reduplicated portion (Duanmu, 2007). This does not happen across all parts of speech but in the context of verbs if it is reduplicated like in the [看看] 'kàn kan' example, the underlying form is /T4 T4/, but on the surface it is realized as [T4 T0]. When it comes to hierarchical organization, this could lead to an environment where tone sandhi and reduplicated tone-dropping *could* occur at the same time (Gao et al., 2021). For instance, if a T3 verb is reduplicated — like [使] 'shǐ' becoming [使使] 'shǐ shi' or 'shì shi' — then in theory it would trigger tone three sandhi *and* also a reduplication transformation. However, they are not triggered simultaneously, there is a hierarchical organization to them where each transformation applies one at a time until the surface form is derived (Lahirir & Marslen-Wilson, 1991).

First however, understanding a simple example of how tone three sandhi occurs in compound verbs will give foundation to the hierarchical organization of transformations in reduplicated verbs.

### 1.2aa Compounds.

Compounds are perhaps the most standard environment for tone three sandhi to occur. Verbs in Chinese may be one morpheme or can be combined with other words to become compound verbs. An example can be seen below:

(1) 领导

| | | |
|---|---|---|
| /lǐng dǎo/ | → | [líng dǎo] |
| /T3 T3/ | → | [T2 T3] |

Collar Guide

'to lead'

In (1) the whole meaning of /领导/ 'lǐng dǎo' is the verb 'to lead' but the exact translation — or gloss — of the word is 'collar guide'. Together, the two are lexicalized into the meaning of, 'to lead'.

As for the tones in (1), it is the exact environment where both morphemes of the verb are tone three /T3 T3/. This then leads to the alternation of the first syllable 'lǐng' (T3) being changed into a T2 for the surface form 'líng dǎo' with the surface tone contour being [T2 T3]. Again, speakers tend to *think* they are pronouncing 领导 as /lǐng dǎo/ with /T3 T3/ pitch contour but in reality, they are pronouncing it as 'língdǎo' with [T2 T3] pitch contour (Duanmu, 2007).

Compounds may be considered more independent as lexical items as their meaning is lexicalized over the entire word. However, this may not be the same for reduplicated verbs.

### 1.2ab Reduplication.

Reduplicated verbs — in the context of this thesis — are derived from a monosyllabic verb such as the previous example of [看] 'kàn' can be reduplicated to become [看看] 'kàn kan' with a diminutive meaning. The difference between compound verbs and reduplicated verbs is that reduplicated verbs are derived from existing monosyllabic verbs in a way that compound verbs are not. Compound verbs are lexicalized for their whole meaning (Duanmu, 2007), meanwhile reduplicated verbs take the meaning of their base and make it the diminutive (Chen et al., 1992). Therefore, these two different kinds of verbs may be processed differently as they are morphologically different.

As for how exactly tone three sandhi may apply reduplicated verbs, the literature is not exactly clear. It is not agreed upon across all researchers whether reduplication creates an environment for tone three sandhi to occur because verb reduplication utilizes a morphological process in Standard Chinese that already has a tone alteration that is not necessarily related to tone three sandhi (Duanmu, 2007; Gao et al., 2021). This alteration is a tone dropping or neutralization of the *second* morpheme of a reduplicated verb making it become a T0.

This verb reduplicated tonal alteration only applies to Chinese monosyllabic verbs. It fully repeats the verb, changing the meaning to its diminutive or a more casual way of saying that verb (Gao et al., 2021). An example of basic reduplication can be seen below:

| | | | | | |
|---|---|---|---|---|---|
| (2) | 学 | → | 学学 | | |
| | [xué] | → | /xué xué/ | → | [xué xue] |
| | [T2] | → | /T2 T2/ | → | [T2 T0] |
| | ‘to study’ | → | ‘to study a little bit’ | | |

In (2), [学] ‘xué’ the non-reduplicated form, the meaning is straightforward as ‘to study’. Meanwhile when it is reduplicated to /学学/ ‘xué xue’ the meaning is changed to ‘to study a little bit’. The reduplicated tone dropping occurs in the surface form of reduplicated verbs where the second morpheme’s tone is dropped (Duanmu, 2007). This means that both morphemes do not carry the same tone. Looking back at (2), notice in the surface form of the reduplicated verb the tone contour is [T2 T0] not /T2 T2/.

However, reduplication has the potential to provide the specific *underlying* form that triggers tone three sandhi /T3 T3/. An example of a possible environment for sandhi is below in (3):

(3) 使使

/shǐ shǐ/

/T3 T3/

‘to try a little bit’

There is debate about which tone alternation comes first to get to the surface form for (3): reduplicated tone dropping or tone three sandhi.

Duanmu (2007) claims that reduplicated tone dropping occurs first and that tone three sandhi is *not* triggered in this environment. If this were the case, the underlying and surface forms for (3) would look like (4) below, where sandhi does *not* occur and the initial morpheme stays the same.

| | | | |
|---|---|---|---|
| (4) | 使使 | → | *使使 |
| | /shǐ shǐ/ | → | *[shǐ shi] |
| | /T3 T3/ | → | *[T3 T0] |
| | ‘to try a little bit’ | | |

Meanwhile Gao et al., (2021) performed a speech perception experiment, treating reduplicated tone three verbs as valid environments for sandhi. They proposed that tone three sandhi occurs *before* the reduplicated tone dropping. Therefore, sandhi *is* triggered in (3)’s environment and proposed (5) as the process to get to the surface form.

| | | | | | |
|---|---|---|---|---|---|
| (5) | 使使 | → | 使使 | → | 使使 |
| | /shǐ shǐ/ | → | [shí shǐ] | → | [shí shi] |
| | /T3 T3/ | → | [T2 T3] | → | [T2 T0] |
| | ‘to try a little bit’ | | | | |

As the literature is unclear on which of these proposals is correct, Gao et al., (2021)’s idea of T3 reduplication being *valid* for sandhi to occur is adopted. While Duanmu (2007), proposes a solid option, his work is mainly theoretical when it comes to his discussion of reduplication and ignoring sandhi in that environment. In the context of a production experiment, following the lead of previous experimental research (such as Gao et al., 2021) is more appropriate. Furthermore, Gao et al., (2021) found no evidence that sandhi *did not* occur in reduplication in their discussion.

While theoretical work is important and another way to approach tone three sandhi, experimentation through perception and production experiments provides quantitative data about the intricacies of this alternation in the minds of natives.

## 1.3 Speech Perception and Production

Speech perception experiments often have results with data that is easy to categorize. For instance, having participants listen to a sentence containing tone three sandhi (manipulating the tone they hear) and ask them to make a comprehensibility judgment with ‘yes’ or ‘no’ keys will give researchers an abundance of easy to decipher data. This could include reaction times and the statistics from the correct answers. Speech perception experiments are vital to the field of linguistics and can find the intricacies of language that may be needed to tease a part more.

Speech production experiments are not as common for tone three sandhi. These types of experiments rarely have binary data that can be easily deciphered. For instance, having participants speak sentences to analyze the way they said a single tone has more factors and labor behind any quantitative findings. Previous researchers have conducted tonal production experiments, Xu (1997) for example, however even they do not include tone sandhi in their results.

Regardless of the lack of binary results, the qualitative data from production experiments is important to further the linguistic field. Elaborating on how exactly people speak and the phenomena that is so heavily referenced (Duanmu 2007, Cheng 1973, etc.), like tone three sandhi, should be tested when new information about it comes out. For instance, the Gao et al., (2021)'s perception experiment found differences between reaction times in processing tone three sandhi between reduplicated and compound verbs. Shifting this experiment into a production experiment could further tease apart a potential spoken difference between these two modalities.

While there is a link between speech perception and production, the details and theories in this field are important when analyzing the results of this thesis. Most research conducted in this area concerns experimenting with second language learners or novel words/sounds. This context is not an exact match for this thesis, however the foundation is a jumping off point for how perception and production are related.

Similar to how Duanmu (2007) noted earlier that most dialect speakers of Chinese only learn enough Standard Chinese to be understood, the same can be said for how most speakers of any language approach production. Baese-Berk & Samuel states that, "... speakers produce in order that perceivers will understand a message," (Baese-Berk & Samuel, 2016, pg 24). Speakers for the most part are not entirely focused on having perfect production. Mitterer & Ernestus (2008) take this even further by saying that the link between perception and production, "... is at an abstract level and governed by phonological relevance," (Mitterer & Ernestus, 2008, pg 173). Therefore, if an aspect of speech has little phonological relevance, it may be dropped or less pronounced.

When it comes specifically to tone three sandhi and its relationship between perception and production there are a few relevant sources (W.Wang and Li, 1967, Zhang & Lai, 2010, Gao et al., 2021) to discuss below.

W.Wang and Li (1967) performed a perception experiment on the Beijing dialect of Chinese and tone three sandhi. They used 130 pairs of disyllabic utterances in Chinese that were either compound tone two tone three (T2 T3) or the environment for sandhi of two tone three's (T3 T3). This list did not discriminate by part of speech or morphology as there were a variety of nouns, verbs, adverbs, and more in their stimulus list. They did not embed these words into sentences nor context, instead they had 14 participants instructed to listen to the isolated stimuli and circle on a worksheet what pair of characters they thought they heard (either T2 T3 compound or T3 T3 compound). In the end all participants got between 49.2 – 54.2% accuracy on correctly identifying all the stimuli, which W.Wang and Li (1967) decided to be not very accurate. Therefore, they concluded that the tone sequence T3 T3 to be homophonous with T2 T3 in neutral speech. This means in the case of speech perception, the way a person hears tone sandhi is the same regardless of whether they heard a production of the underlying or surface form.

Furthermore, in the previously discussed production study conducted by Zhang and Lai, (2010) they compared the production of two types of sandhi: tone three sandhi and half third sandhi. They tested participants pronunciation of these sandhi's in both real and 'wug' words by having them listen to two monosyllabic utterances and asked them to say them together as a disyllabic utterance. Every stimulus's initial utterance was a tone three (T3), and the second one could have been any tone of Chinese (T1, T2, T3, or T4). They hypothesized that on 'wug' words, the half-third sandhi would be more productive than tone three sandhi (in T3 T3 combinations) due to half-third sandhi having more phonetic motivation. However they found the opposite, that tone three sandhi was actually more productive in 'wug' words than real words citing it possible that, "the production of the third-tone sandhi is influenced by a greater perceptual perseveration effect from the input than that of the half-third sandhi," (Zhang & Lai, 2010, pg 189). Essentially this means that after participants hear two full tone three's [T3] [T3] may 'prime' them into pronouncing with tone three sandhi (surface from becoming /T2 T3/). Yet Zhang and Lai, (2010) explains that the theory of tone alone being a prime is controversial which makes this conclusion weak. Still, Zhang and Lai, (2010) shows the unique relationship between perceiving tones and pronouncing them in the context of tone three sandhi. Where participants will hear one thing, are asked to repeat it, and pronounce the tonal contours differently than what they heard. Therefore this thesis proposes diving into specific parts of speech that have also been found to be perceived differently (Gao et al., 2021) and further see if perhaps this difference has subtler influences on production.

One such difference in perceiving different parts of speech containing tone three sandhi is in Gao et al., (2021)'s perception experiment. They found participants taking more time to reject or accept tone three sandhi in reduplicated and compound verb environments. With the knowledge of how the relationship between tone three sandhi in real and fake words perceiving and production are different (Zhang & Lai, 2010), continuing this line of research by analyzing tone three sandhi in the context of real verbs in Standard Chinese that are perceived differently (Gao et al., 2021) will further the field of comparing perception and production in linguistics. Since, in general, it is concluded that, "the relationship between perception and production is more complex than previous studies have suggested," (Baese-Berk & Samuel, 2016, p. 34). Therefore, converting Gao et al., (2021)'s perception experiment into production could add to the discussion and lend weight to whether or not the relationship is linked in the context of native Standard Chinese speakers.

### *1.3a Previous Experiment*

Gao et al., (2021)'s conducted a perception experiment concerning certain environments for tone three sandhi. Their method was to show native Standard Chinese speakers a full sentence but two characters at a time. In these sentences, there were target characters highlighted in red — usually related to sandhi. An example of how this could have looked is outlined below:

**Figure 2**

*Gao et al., (2021)'s potential visuals*

| screen 1 | screen 2 | screen 3 | screen 4 | screen 5 |
| --- | --- | --- | --- | --- |
| 这个<br>（zhège) | 工具<br>(gōngjù) | 请你<br>（qǐng nǐ） | 稍微<br>(shāowéi) | 使使<br>(shǐ shǐ) |

*Note* (translation): 这个 工具 请你 稍微 使使

/zhège gōngjù qǐng nǐ shāowéi shǐ shǐ/

'please try this tool a little bit.'

When these red characters appeared on the participants screen, an audio file would play. These audios were supposed to be the verbal version of the written characters. The only difference was that the tones were different. The participants were instructed to identify whether or not the audio matched the red-coded characters.

These characters were able to be, "... legitimately combined with T1, T2, and T3," (Gao et al., 2021, pg 6) as they were testing many combinations of tones on these target verbs however their results, and for the purposes of this thesis, the only significant findings were in tone two and tone three combinations.

After collecting reaction times and responses from participants with the aforementioned methodology, the results showed that, "... responses given to the underlying tones were significantly different between T2 [reduplication] and T3 [compounds where sandhi occurs] (...) but not between any other two constructions." (Gao et al., 2021, pg 8). This means that when participants heard the underlying tones for T3 compounds — the category in which sandhi takes place — it was not the same response as they would have given for the control of T2 reduplication — where sandhi does not occur. This was an expected result as there is potential confusion due to the difference in underlying and surface forms in T3 compounds related to sandhi that is not present in T2 reduplicated verbs.

Furthermore, there was not a significant difference between the reaction times of the two sandhi stimuli categories (T3 reduplication and T3 compounds). However, participants were, "more inclined to reject the T2 [the surface tone] in reduplication than in compounds," (Gao et al., 2021, pg 9). Which suggests that, "the underlying and the surface representations were more likely to compete with each other in lexical compounds, as opposed to reduplication," (Gao et al., 2021, pg 9).

As a whole, these sandhi categories had two main results:

1. Rejecting the underlying T3 was harder than accepting it
2. Participants were confused at parsing the sandhi syllable as its surface T2

(Gao et al., 2021, pg 11)

This means that in native speakers' minds, the difference between the underlying and surface form is not clear when it comes to speech processing. Participants will hear the underlying form of these constructions /T3 T3/ and have a harder time rejecting it than when they hear [T2 T3].

This brings up the question that maybe this confusion comes from the way sandhi is pronounced natively. Perhaps the reason native speakers have a hard time with this is because they are pronounced differently. Recall the initial findings of this experiment where they, "observed an easier access to the surface tone in sandhi compounds than in sandhi reduplication," (Gao et al., 2021, pg 11) which suggests that the surface form tone mapping plays a bigger role in compound verbs than in reduplicated verbs. Therefore, perhaps there is a difference in tone pronunciation between these two categories.

Regardless of tone, they conclude saying that their results, "show that while the underlying tone representation was strongly activated, the surface tone representation was relatively weak and less likely to be accepted," (Gao et al., 2021, pg 14). This means that in tone three sandhi categories (i.e. T3 compounds and T3 reduplicated verbs) participants, generally, reacted quicker when hearing a T3 audio than a T2 audio. When they were presented with a tone three sandhi category but heard a T2 audio, they were less likely to accept it as matching or took a longer time saying that it matched.

## 1.4 The Present Study

There are a variety of reasons Gao et al., (2021) could have found these results. Previous studies have shown similar findings such as the perception study of W.Wang and Li (1967) that showed that listeners did not find a distinguishable difference when T3 changes to T2 in tone sandhi environments.

One possibility is that these two categories of sandhi (compound and reduplicated verbs) may be pronounced differently. There have been studies that look at tone sandhi across a variety of parts of speech (W.Wang and Li 1967, Nixon et al., 2015, etc.) or even fake words (Zhang et al. 2015, Zhang & Lai, 2010, etc.) without distinguishing between morphology or without embedding the stimuli in context. Few have tested specific morphological structures and parts of speech like Gao et al., (2021) in relation to tone three sandhi production. Therefore, given the findings in Gao et al., (2021)'s perception experiment, which found, "easier access to the surface tone in sandhi compounds than in sandhi reduplication…" (Gao et al., 2021, pg 11), perhaps the pronunciation of sandhi in compounds is more like the expected surface of a tone two (T2) than it is in reduplication.

My research question is twofold:

1. Is tone three sandhi pronounced differently in reduplicated and compound verb constructions?
2. Do produced tone contours reflect what has been found from perception experiments?

With the discussion already given on these topics, there are some predictions for this thesis. Given the different morphological constructions between compound and reduplicated verbs in Chinese and the previous findings from perception experiments, I do anticipate a different production of sandhi. More specifically, I predict that the initial morpheme in sandhi compounds will be more like a standard T2 than the initial morpheme in sandhi reduplicated verbs, which will be more like T3 — lower and perhaps with a slight dipping motion.

When I compare the tone contours of the initial morpheme of sandhi compounds and reduplicated verbs in comparison with some T2 controls, I expect the sandhi compounds to match more with the controls than the sandhi reduplicated verbs. A visualization of this hypothesis is in the Table 2 below:

**Table 2**

*Hypothesized Breakdown of Majorities*

| | **Strong T2 Contour** | **Weak T2 Contour** |
|---|---|---|
| **T3 compound verbs**<br>/T3 T3/ → [T2 T3] | X | |
| **T3 reduplicated verbs**<br>/T3 T3/ → [T2 T0] | | X |
| **T2 controls** | X | |

In Table 2, I hypothesize that the majority of contours of stimuli will be in the boxes where I placed an ‘X’. Therefore the majority of T3 compound verbs should be in the cluster that has a ‘strong T2 contour’ along with the T2 controls. A ‘strong T2 contour’ would be a contour that looks like T2 (from Figure 1), where the pitch contour starts low and ends high). Meanwhile, the majority of T3 reduplicated verbs should be in the cluster that has a relatively ‘weaker T2 contour’ meaning that perhaps the pitch contour starts higher than the ‘strong T2 contour’ or where it dips down slightly after the onset before rising — similar to a T3.

# 2 Methodology

To test this hypothesis, I shifted Gao et al., (2021)’s perception experiment into a production one. Through a remote experiment, participants were presented with sentences (similar to Gao et al., 2021’s stimuli) and asked to read them aloud. Their responses were recorded and collected for later analysis.

## 2.1 Participants

The participants for this study were four self-identified native Standard Chinese (or Mandarin) speakers. One person identified with two native languages: Mandarin and Cantonese. All participants were at least bilingual in English and Mandarin along with identifying proficiencies in many other languages, including other Chinese dialects: Cantonese and Hokkien.

To confirm participants' language background, I asked about their geographical background. Asking where participants grew up and what areas they had spent a significant amount of time in as it may influence how they speak. As discussed earlier, Standard Chinese is based on Mandarin, a northern dialect. Therefore, I only accepted participants that identified as a native Standard Chinese speaker *and* spent a significant amount of time in either the north of China or a place where the official regional dialect was Mandarin.

The participants identified with growing up in a variety of places. Many identified with spending a significant amount of time in regions in the south of China that spoke Cantonese. However, they specified that they lived in southern subregions whose official language is Mandarin (Shenzhen and Zhuhai). Only one participant identified with growing up in the north of China.

Furthermore, at the time of the experiment, all participants were living in the Netherlands. They had lived there for as little as 7 months to 10 years. All participants were between the ages of 22 and 30, identifying their education level as master students.

## 2.2 Stimuli

To preserve as much of the original processing experiment as possible, the target stimuli was taken from the previous Gao et al., (2021) experiment's appendix along with the carrier sentences. However, I added a new control category entitled 'control tone two three compounds' (T2 T3 COM) to control for the previous T3 COM category.

Overall, there were four categories of stimuli:

1. Tone three sandhi in compounds (T3 COM)
2. Tone three sandhi in reduplication (T3 RED)
3. Control tone two reduplication (T2 RED)
4. Control tone two three compounds (T2 T3 COM)

### *2.2a T3 COM*

In the first category, 'tone three sandhi in compounds' (T3 COM), this is a classic environment for tone three sandhi to occur. Here is another example of one of the target words:

(6) 引导

| | | |
|---|---|---|
| /yǐn dǎo/ | → | [yín dǎo] |
| /T3 T3/ | → | [T2 T3] |
| ‘to guide’ | | |

The underlying form /T3 T3/ is different from the surface form [T2 T3]. This has been shown in the previous processing experiment (Gao et al., 2021) that participants have an *easier* time accessing the surface form [T2 T3] than in other environments for tone three sandhi (e.g. T3 RED).

The purpose of this stimulus category is to test if the reason the surface form is more accessible is because the pronunciation is exactly as the phonological rules suggest — the initial morpheme is a strong tone two contour.

### *2.2b T3 RED*

In the second category, ‘tone three sandhi in reduplication’ (T3 RED), this is another example of an environment for tone three sandhi to occur. Another example of one of the target words is below:

(7) 想想

| | | | | |
|---|---|---|---|---|
| /xiǎng xiǎng/ | → | [xiáng xiǎng] | → | [xiáng xiang] |
| /T3 T3/ | → | [T2 T3] | → | [T2 T0] |
| ‘to think for a little bit’ | | | | |

The underlying form /T3 T3/ is different from the surface form [T2 T0]. In the previous processing experiment (Gao et al., 2021), participants had a *harder* time accessing the surface form [T2 T0] than in other environments for tone three sandhi (e.g. T3 COM).

The purpose of this stimulus category is to test if the reason the surface form is less accessible is because the pronunciation is different than what the phonological rules suggest. The initial morpheme tonal contour may be a weaker tone two contour than in other categories where the surface form is more accessible (e.g. T3 COM).

### *2.2c T2 RED*

In the third category, ‘control tone two reduplication’ (T2 RED), this is *not* an example of a tone three sandhi environment. It serves as a control category to see a true T2 contour in comparison to tone sandhi T2 contour.

An example of this control category is below:

(8) 谈谈

/tán tán/ → [tán tan]

/T2 T2/ → [T2 T0]

‘to talk’

The only transformation that occurs here is tone dropping/neutralization that is expected in surface forms of reduplicated verbs.

The T2 RED category was used as a control in the previous processing experiment (Gao et al., 2021). This is not a place where tone sandhi takes place; however, the underlying form /T2 T2/ is still different from the surface form [T2 T0] due to tone neutralization.

For the purposes of this production experiment, the tonal contours from the initial morpheme (a tone two) will be used as a control for the category of T3 RED. If tone sandhi occurs perfectly, the tonal contours should be similar.

### *2.2d T2T3 COM*

The category T2 T3 COM is the novel control category in this production experiment. This is not a place where tone sandhi takes place nor an environment for tone neutralization (e.g. the reduplication categories). The surface form and underlying form are exactly the same, as seen below:

(9) 来访

/lái fǎng/ → [lái fǎng]

/T2 T3/ → [T2 T3]

‘to visit’

Furthermore, this tonal category is the expected surface tonal pattern of T3 COM. Therefore, this category serves as a reference for how sharp — or not — the initial morpheme in T3 COM production is.

## 2.3 Procedure

The production experiment was programmed using Gorilla software (www.gorilla.sc). Due to concerns of the experiment being too long, the target stimuli were shortened from 15 target words/carrier sentences per category (used in Gao et al., 2021) to 10 sentences per category. Similarly, the filler stimuli were significantly cut down from 100 filler sentences (used in Gao et al., 2021) to only 10. Participants averaged 16 minutes to complete the entire experiment.

The experiment was conducted entirely in Standard Chinese. It opened with a short questionnaire asking for language background, geographical background, and age. If participants did not fit the desired criteria, the programming would exit them from the experiment.

Furthermore, the experiment was designed specifically so participants would have to click as little buttons as possible — as it could mess with the audio with loud clicks.

Participants then were shown the task instructions, explaining the process of the experiment. A translation of the instructions is provided below:

1. Fixate on the cross in the middle of the screen
2. A sentence will appear after a few seconds
3. You will have 7 seconds to read aloud the sentence before it disappears
4. Repeat

They were then asked to consent to the experiment and for their data to be used for research purposes.

Once participants consented, they were given three trial sentences to get them used to the pace of the stimuli appearing and disappearing. A countdown timer was shown above the sentence (counting down from 7 to 0) along with a progress bar to notify participants how far along into the experiment they were. A visual of the sequence of trial screens participant saw are below:

**Figure 3**

*Visual of Experimental Trial*


Notice how this is different from the processing experiment conducted by Gao et al., (2021). They highlighted their target words in red, meanwhile my experiment presents all the text in plain black text. This is an attempt for participants to not know what the target words are so they will not overemphasize any one part of the sentence in an attempt to provide better sound bites.

After the trial runs, participants then moved onto the bulk of the experiment. They read aloud a total of fifty Standard Chinese sentences, with a break halfway through. Due to the automatic nature of the experiment, the change in duration was due to how long or little the participant took on their break.

# 3 Results

The bulk of analysis conducted used f0 contour clustering (Kaland, 2021). This application gathers pitch contours from many sound files and combines them into groups based on similar contours. It also provides a visual of these contours that can show the difference between the categories. This will be used to identify the tone patterns to prove or disprove the hypothesis that: T3 COM will have a stronger and more recognizable tone two than T3 RED.

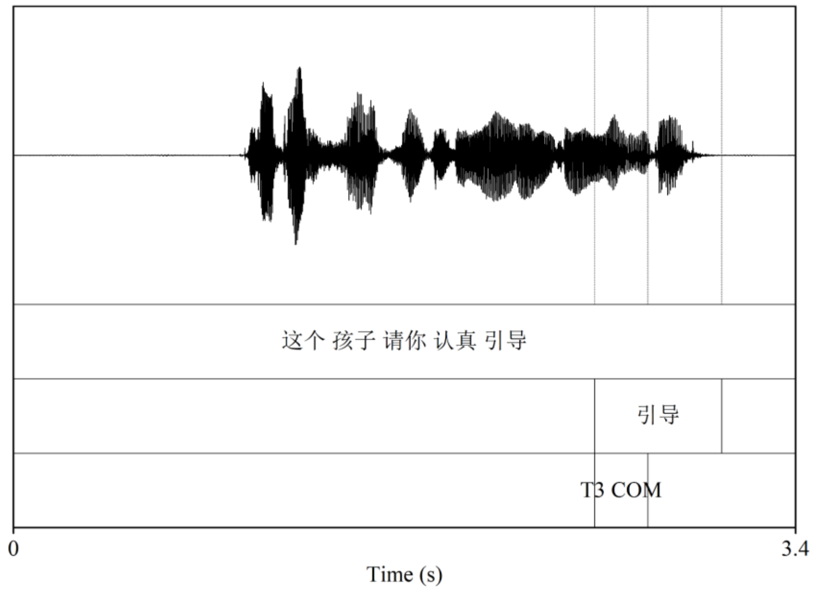
## 3.1 Data Pre-Processing

The sound files were downloaded from gorilla as webM files and were converted into .wav files in Audacity (v3.4.2: Audacity®) using macro manager tools. The trial sentences and the filler sentence sound files were discarded here as they are not needed for the analysis. An R (v4.3.3; R Core Team 2024) code ran on the folder containing the sound files to give them comprehensive names (e.g. which order it was presented to the participant). The code incorporated an excel sheet that had all the stimulus names and categories (T3 COM, T3 RED, T2 RED, etc.). Along with the renamed sounds, the R code also produced a .csv file holding the name of the file and which stimuli sentence that was being said in it.

With the sound files and the .csv file, Praat (Boersma and Weenink, 2024) scripts created text grids for each file and filled them with the sentence that was being said. Then one by one, I checked each sound file, creating tier intervals that contained the desired morphemes — the initial morpheme of all sandhi categories. An example of how this would look is in Figure 4 below:

**Figure 4**

*Praat Oscillogram and Textgrid with Three Tiers*




*Note* (translation): 这个 孩子 请你 认真 引导

/zhège háizi qǐng nǐ rènzhēn yǐndǎo/

'please guide this kid carefully'

Here you can see the oscillogram in the top visual, below that is the corresponding sentence the participant spoke. In this case, the sentence was "这个 孩子 请你 认真 引导 (zhège háizi qǐng nǐ rènzhēn yǐndǎo)" which translates to "Please guide this kid carefully". Like the methodology section said, the final two characters was the target word, in this case this was the polysyllabic verb "引导 (yǐndǎo)" which was from the T3 COM stimuli. In the second tier from the bottom, you will see the entire compound word was narrowed down from the sentence: "引导 (yǐndǎo)". And in the bottom row, this is where just the first morpheme "引 (yǐn)" was specified as, "T3 COM" as a reminder and for later analysis when plugged into clustering software. This was the target morpheme and what was later plugged into the contour clustering software.

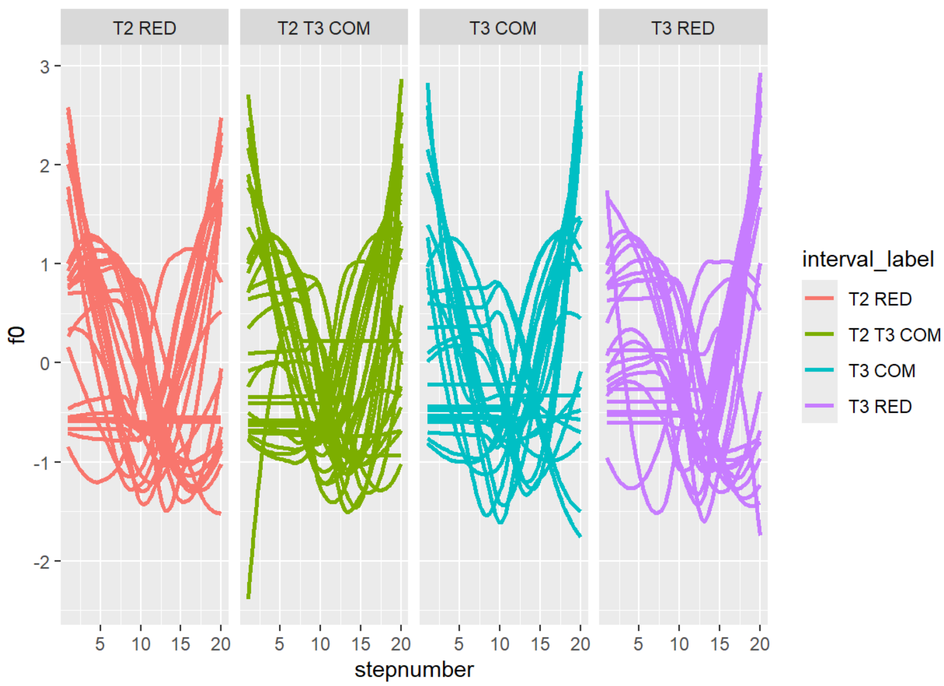
The contour clustering application used the sound files and text grids to produce a dataset of the f0's — pitch points — in the desired tier intervals (in this case the initial morpheme of the target verb). The pitch floor was kept at 75 Hz and the ceiling at 500Hz with the timestep being 10ms. As for the f0 (pitch) fit — measuring how well the f0 matches and selecting good candidates for the data — was changed from relatively weak 0.52 to a more strict 0.8 to gather the best measurements from the data. The number of measurement points was maintained at the default 20, essentially meaning that this was the number of points across the contour I will be taking. The smoothing bandwidth was also kept at the default 1, making sure that the sometimes-faulty measurements could be smoothed over in case there was an issue with the participants microphone or background noise.

This application took the 20 measurements of pitch across the desired sound bite to place in a table as a dataset. The software standardized for multiple speakers, normalized f0 measures, and cleaned out the data for any f0's measurements that were still unclear even after the adjustments described above. This resulted in only 104 sound files remaining out of the initial 159 — originally 160 but one audio file from one speaker was consistently showing as an outlying contour pattern and was removed from the dataset.

From this and some short R codes, I was able to visualize my data by sentence and by type of target word that was being pronounced. This can be shown in the Figure 5 below:

**Figure 5**

*Smoothed (normalized f0) Curves of Pitch Contours in Production Task using the geom smooth Function*



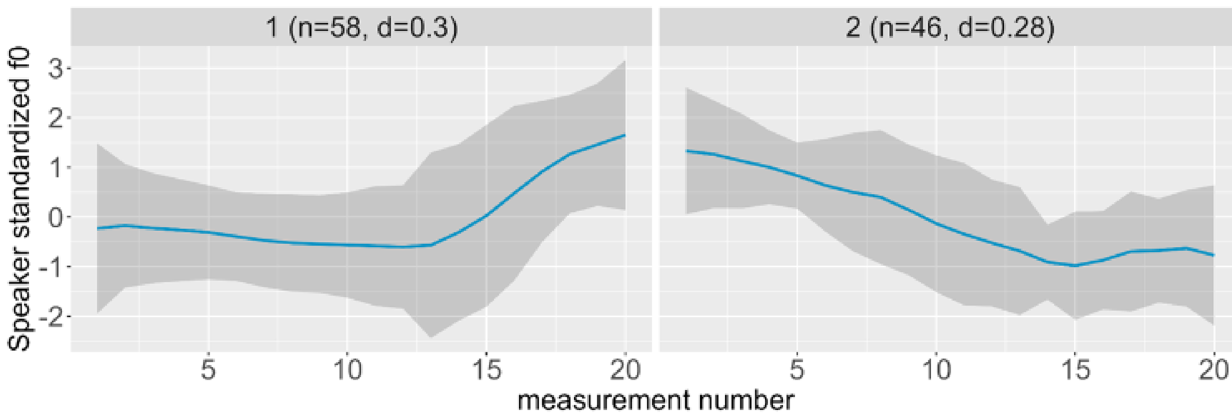

The figure is a bit overwhelming at first, however you can see four graphs, one for each category of stimuli (T2 RED, T2 T3 COM, T3 COM, and T3 RED). Each line in these graphs is each individual target contour for all the stimuli. These are the measurements that were normalized across all 20 pitch measurements. Based on this figure alone, there is not much that can be deduced. Due to the chaotic nature of the measurements, these need to be clustered together with more adept software to handle all the possible groupings and patterns in the contours. This can be done in Kaland (2021)'s contour clustering software.

## 3.2 Contour Clustering

The contour clustering software used the same dataset used above and was set to two clusters. All of the stimuli should fit into only one category and look like a T2. However, the research question is to see if there is a difference between T3 COM and T3 RED pronunciation of this T2, therefore with two clusters the majority of sound files should fit into one category and a smaller portion should fit into the other if there *is* a difference in how T3 COM and T3 RED are produced.

**Figure 6**

*Contour Clustering Output of Initial Dataset*




In Figure 6, there are two significant groupings of the data where the contours were similar to each other (Cluster 1 and Cluster 2). Across the x axis, there are the 20 measurements taken from each sound bite lined up with the y axis being the standardized pitch.

Remembering Figure 1, with all the tone patterns for tones in Standard Chinese, these two images above (in Figure 6) show some similar patterns. In the Cluster 1 (seen in Figure 6), this looks like a contour between a Tone 2 and a Tone 3. Where it starts middle ground, falls just slightly then rises.

In Cluster 2, the contour looks more like a falling tone. It starts relatively high and falls, like a Tone 4. While the second image is not an expected result from this production experiment, it is interesting to see what category of stimuli falls into each of these clusters.

Below is a table of which category of stimuli the software grouped into each cluster:

**Table 3**

*Contour Clustering Output of Initial Dataset Breakdown*

| | Cluster | | |
|---|---|---|---|
| | **1**<br>**T2 Contour** | **2**<br>**T4 Contour** | **Total** |
| **T3 COM** | 15 | 9 | 24 |
| **T3 RED** | 14 | 10 | 24 |
| **T2 RED** | 12 | 15 | 27 |
| **T2 T3 COM** | 17 | 12 | 29 |
| **Total** | 58 | 46 | 104 |

For the two clusters, the majority of all categories except one fell into Cluster 1. The only category that had the majority of their stimuli in Cluster 2 was T2 RED. Now, there is still a lot of overlap between the two clusters, and a significance test could be helpful to see if this pattern is significant, but as it stands, 15 out of 27 T2 RED's fell into the second category (while only 12 fell into the first).

For the rest of the categories of stimuli (T3 COM, T3 RED, and T2 T3 COM) they had their majority in Cluster 1.

While this is a breakdown of the overall data, perhaps taking a closer look into each of these clusters and if they can be broken down even more, may be beneficial to fully fleshing out this analysis.

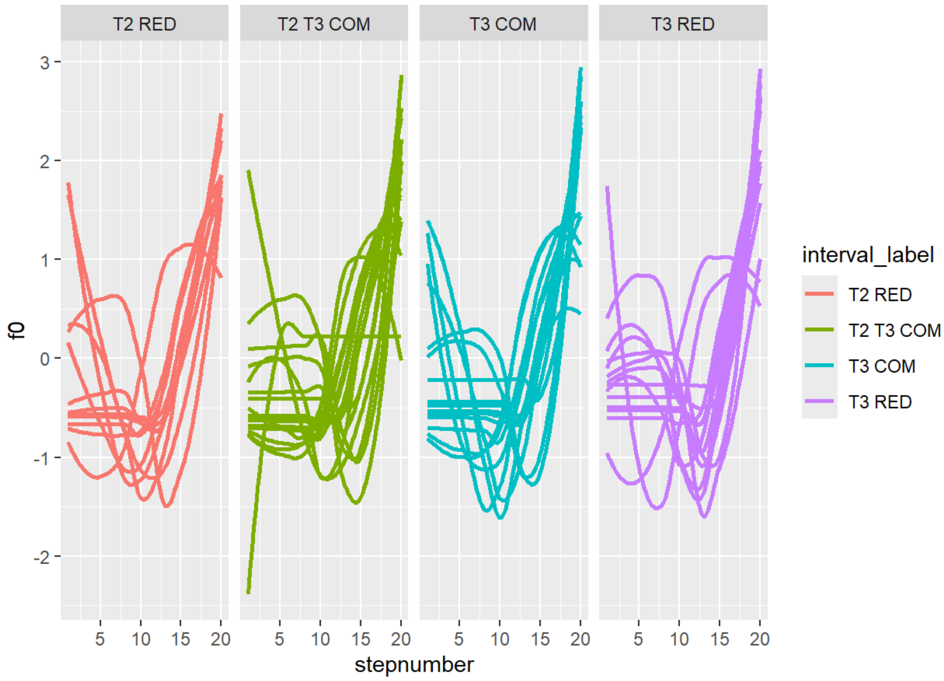
After manually separating the data that fell into Cluster 1 from Cluster 2, I created two new datasets that could be plugged into the contour clustering software. The pre-processing steps were not changed, and the same process was run again.

## 3.3 Within Cluster 1

Cluster 1 contained 58 pitch contours that ranged over all the categories of stimuli. For a quick look at what these contours looked like, I reran the R code that gave me plots for the data initially. The following figure shows the pre-cluster data of *just* Cluster 1.

**Figure 7**

*Smoothed (normalized f0) Curves of Cluster 1 Pitch Contours in Production Task using the geom smooth Function*

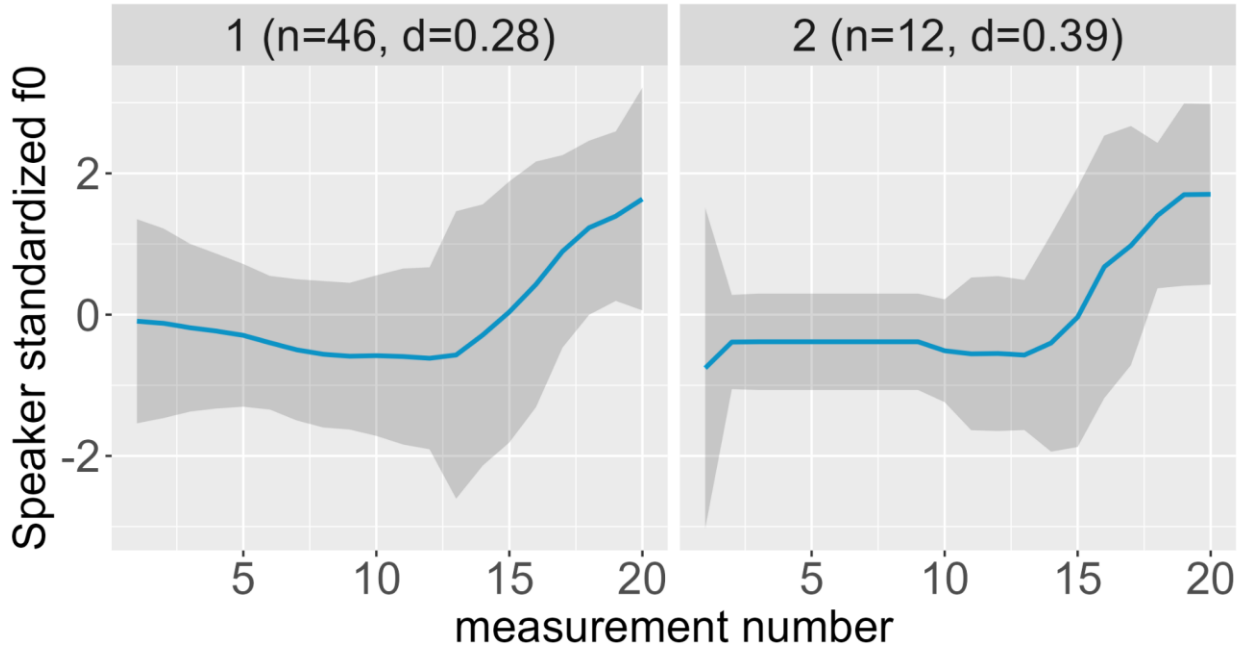



In Figure 7, the data does seem a bit cleaner as it is just a little over half of the original dataset (58 out of the initial 104). Nothing can be determined just by looking at this, however in comparison to Figure 5, this is already looking to be a cleaner dataset.

Therefore, this dataset was plugged back into the contour clustering software to see how they would combine the data this time. The following figure shows the data from just Cluster 1:

**Figure 8**

*Contour Clustering Output of Cluster 1*




When setting the amount of clusters to two, these are the two images (Cluster 1A on the left and Cluster 1B on the right) the software provided. And at first glance, this does seem to be more of what was hypothesized. These two are more of the Tone 2 images that were predicted, with Cluster 1B looking like a stronger T2 than the initial one — as Cluster 1A has more of a slower fall and rise, similar to a T3.

Below is a table of which category of stimuli the software grouped into each cluster:

**Table 4**

*Contour Clustering Output of Cluster 1 Breakdown*

| | Cluster | | |
|---|---|---|---|
| | **1A** Weak T2 | **1B** Strong T2 | **Total** |
| **T3 COM** | 11 | 4 | 15 |
| **T3 RED** | 13 | 1 | 14 |
| **T2 RED** | 10 | 2 | 12 |
| **T2 T3 COM** | 12 | 5 | 17 |
| **Total** | 46 | 12 | 58 |

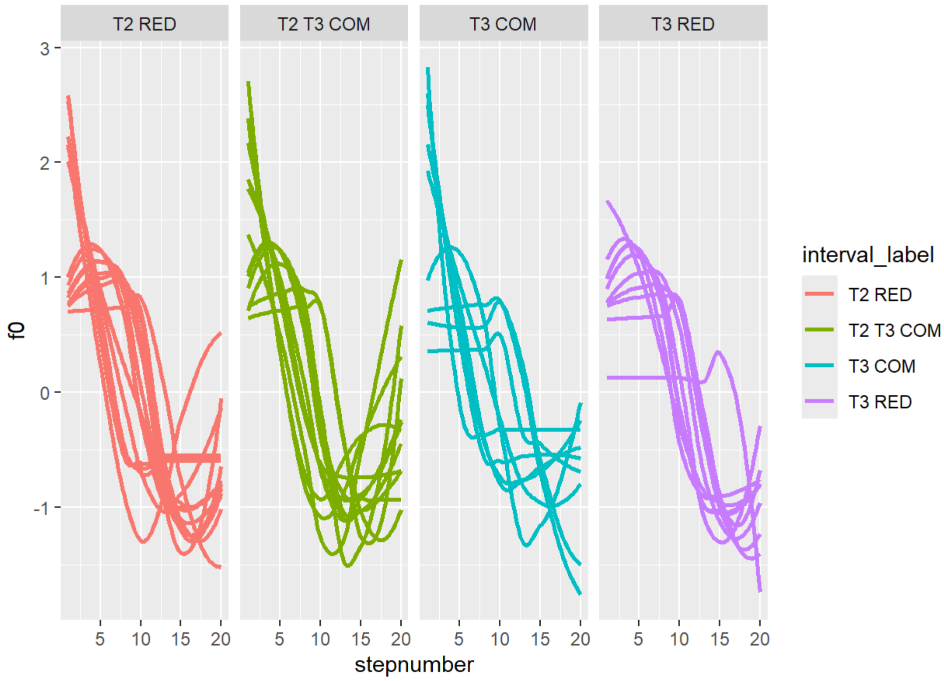
In Table 4, the majority of all categories fall into the Cluster 1A. Cluster 1A has a total of 46 contours (out of 58) while Cluster 1B only has 12. However, within Cluster 1B, the majority of the 12 come from the T2 T3 COM and T3 COM stimuli categories.

## 3.4 Within Cluster 2

Cluster 2 contained 46 pitch contours that ranged over all the categories of stimuli. For a quick look at what these contours looked like, I reran the R code that gave me plots for the data initially. The following figure shows the pre-cluster data of *just* Cluster 2.

**Figure 9**

*Smoothed (normalized f0) Curves of Cluster 2 Pitch Contours in Production Task using the geom smooth Function*

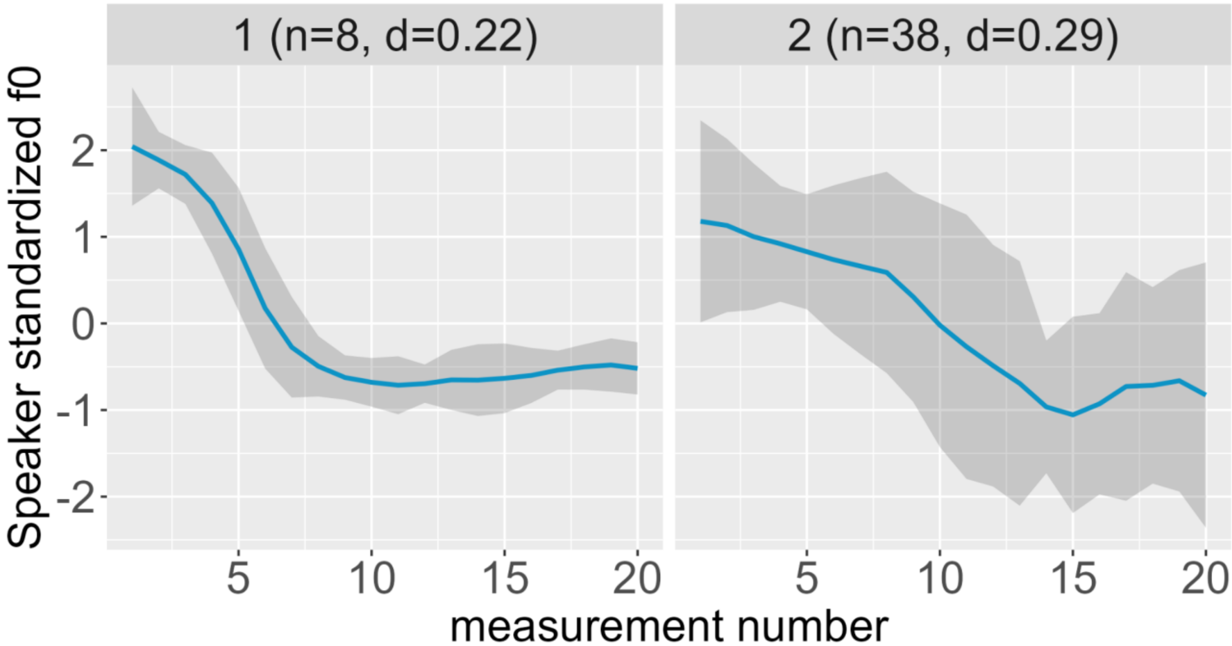



In Figure 9, the data does seem a bit cleaner as it is just a little less than half of the original dataset (46 out of the initial 104). Nothing can be determined just by looking at this, however in comparison to Figure 5, this does look to be a cleaner dataset.

Therefore, this dataset was plugged back into the contour clustering software to see how they would combine the data this time. The following figure shows the data from just Cluster 2:

**Figure 10**

*Contour Clustering Output of Cluster 2*




When setting the number of clusters to two, these are the two images (Cluster 2A on the left and Cluster 2B on the right) the software provided. Cluster 2 was the cluster with the unexpected tone four (T4) contour, and within it itself, it seems that they are still T4's contours with one being a stronger T4 (Cluster 2A) with a sharp downturn than the other relatively weaker T4 (Cluster 2B) that has a more gradual fall.

Below is a table of which category of stimuli the software grouped into each cluster:

**Table 5**

*Contour Clustering Output of Cluster 2 Breakdown*

| | Cluster | | |
|---|---|---|---|
| | **2A**<br>**Strong T4** | **2B**<br>**Weak T4** | **Total** |
| **T3 COM** | 3 | 6 | 9 |
| **T3 RED** | 0 | 10 | 10 |
| **T2 RED** | 2 | 13 | 15 |
| **T2 T3 COM** | 3 | 9 | 12 |
| **Total** | 8 | 38 | 46 |

In Table 5, the majority of all categories fall into the Cluster 2B. Cluster 2B has a total of 38 contours (out of 46) while Cluster 2A only has 8. However, within Cluster 2A, the majority of the 8 come from, the T2 T3 COM and T3 COM stimuli categories. Meanwhile the reduplicated stimuli categories (T3 RED and T2 RED) take up the majority of Cluster 2B.

# 4 Discussion

In summary, for the two initial clusters — using the maximum amount of data — the majority of all categories except one fell into Cluster 1 (with the expected T2 contour). The only category that had the majority of their stimuli in Cluster 2 (with the unexpected T4 contour) was T2 RED. Meanwhile the majority of T3 COM, T3 RED, and T2 T3 COM stimuli fell into Cluster 1.

When breaking down the categories even further, Cluster 1 resulted in the expected T2 looking clusters, meanwhile Cluster 2 continued to show the unexpected T4 clusters. Within both of these sub clusters, the split between them was more uneven than the initial one. Cluster 1 was split between Cluster 1A (weak T2) that had 46 contours within it and Cluster 1B (strong T2) the other containing 12; meanwhile Cluster 2 was split between Cluster 2A (strong T4) having 8 contours within it and Cluster 2B (weak T4) having 38. There did not seem to be a meaningful distribution of specific stimuli categories within these sub clusters.

With these findings, we can readdress the research questions of this thesis:

1. Is tone three sandhi pronounced differently in reduplicated and compound verb constructions?
2. Do produced tone contours reflect what has been found from perception experiments?

## 4.1 Answering RQ1: Is tone three sandhi pronounced differently in reduplicated and compound verb constructions?

The importance of understanding if tone three sandhi is pronounced differently in reduplicated and compound verb constructions is due, not only because of the different morphological constructions, but also because of the varying underlying and surface forms. The way people think about language is not always the same as how people actually speak it. This is known as a difference between underlying forms and surface forms.

When it comes to tone three sandhi and the morphological structures of Chinese that contain those environments, the differences between these morphological structures may lead to a different pronunciation. Tone three compound verbs undergo just one transformation before being pronounced — sandhi — meanwhile reduplicated verbs have multiple (Gao et al., 2021).

Tone three reduplicated verbs not only trigger tone three sandhi, but they also trigger reduplicated tone neutralization. In most cases with reduplicated verbs, in the surface form, the second tone is dropped. For instance an underlying /T2 T2/ will be realized and pronounced on the surface as [T2 T0]. However, due to the knowledge that transformations like this are more of a hierarchical organization (Lahirir & Marslen-Wilson, 1991) it was proposed to follow Gao et al., (2021)'s proposal of tone sandhi occurring — if applicable — *before* reduplicated tone dropping. Therefore, in an underlying form like /T3 T3/ the first transformation will be sandhi resulting in a medial form like [T2 T3] before reaching the surface form [T2 T0].

Now, due to these varying methods in deriving surface forms — compound sandhi verbs merely undergoing one transformation while reduplicated sandhi verbs undergoing two — will this affect how the tone contours are pronounced on the surface? From the given data, the answer is a tentative 'no'.

It was hypothesized earlier that when comparing production of T3 COM and T3 RED, T3 COM might have a sharper T2 than T3 RED. Recall Table 2:

**Table 2**

*Hypothesized Breakdown of Majorities*

| | **Strong T2 Contour** | **Weak T2 Contour** |
|---|---|---|
| **T3 compound verbs**<br>/T3 T3/ → [T2 T3] | X | |
| **T3 reduplicated verbs**<br>/T3 T3/ → [T2 T0] | | X |
| **T2 controls** | X | |

This table displayed where I hypothesized the majorities of the contours would fall into. This was specifically showing that T3 COM and T3 RED were in different categories. However, this is not where the majorities actually fell when looking at all of the data. Here is a table displaying where the majorities of each categories stimuli actually fell:

**Table 6**

*Hypothesized and Contour Clustering Output Breakdown of Majorities*

| | **Hypothesized**<br>**A Strong T2 Contour** | **Hypothesized**<br>**A Weak T2 Contour** | **Cluster 1**<br>**A T2 Contour** | **Cluster 2**<br>**A T4 Contour** |
|---|---|---|---|---|
| **T3 compound verbs**<br>/T3 T3/ → [T2 T3] | X | | X | |
| **T3 reduplicated verbs**<br>/T3 T3/ → [T2 T0] | | X | X | |
| **T2 reduplicated verbs**<br>/T2 T2/ → [T2 T0] | X | | | X |
| **T2 T3 compound verbs**<br>/T2 T3/ → [T2 T3] | X | | X | |

As the research question asks about the differences between sandhi in compounds and reduplicated verbs, likewise, the answer should lie in the stimuli categories T3 COM and T3 RED. Recall that in the initial Contour Clusters 1 and 2, T3 COM and T3 RED had nearly an identical ratio between the two clusters.

Table 3 showed that where T3 COM had 15 contours fitting Cluster 1 (the T2 contour) and 9 fitting Cluster 2 (the T4 contour). Similarly, T3 RED had 14 contours fitting Cluster 1 (the T2 contour) and 10 fitting Cluster 2 (the T4 contour). With the initial dataset, there does not seem to be a clear difference between the two categories as the previous perception experiment led me to hypothesize.

Keeping in mind that the majority gap was not large between any of the groups in this dataset — and also that there was an unexpected T4 contour (Cluster 2) in the results — more results could still be derived.

When looking deeper into the expected T2 contour cluster (Cluster 1), the majority breakdown looks like Table 7 below:

**Table 7**

*Hypothesized and Contour Clustering Output of Cluster 1 Breakdown of Majorities*

| | **Hypothesized**<br>**A Strong T2 contour** | **Hypothesized**<br>**A Weak T2 contour** | **Cluster 1A**<br>**A Weak T2 contour** | **Cluster 1B**<br>**A Strong T2 contour** |
|---|---|---|---|---|
| **T3 compound verbs**<br>/T3 T3/ → [T2 T3] | X | | X | |
| **T3 reduplicated verbs**<br>/T3 T3/ → [T2 T0] | | X | X | |
| **T2 reduplicated verbs**<br>/T2 T2/ → [T2 T0] | X | | X | |
| **T2 T3 compound verbs**<br>/T2 T3/ → [T2 T3] | X | | X | |

In Table 7, this cluster had two contours within it: Cluster 1A that was more of a mix between T2 and T3 (labeled 'weaker T2') with a slight dip before rising and Cluster 1B was a standard T2 with no major dip. However, all of the stimuli category majorities fell into Cluster 1A with the weaker T2.

Therefore, given the results of where the majorities fell in the initial dataset and the deeper dive into Cluster 1, there is not a major difference between how T3 COM and T3 RED sandhi are pronounced.

What this means in terms of answering the research question on a broader scale, with the given data and derived results, is that it does not seem that the varying transformations between underlying and surface forms have a major effect on how the surface forms tonal contours are pronounced. Although T3 RED undergoes more transformations from the underlying to surface than T3 COM, their surface tonal contours are not distinguishable in any major way with this data.

## 4.2 Answering RQ2: Do produced tone contours reflect what has been found from perception experiments?

Converting a perception experiment into a production experiment gives this thesis a unique opportunity of comparing specific results found in the perception experiment nearly exactly. With the same stimuli, similar procedure, and using their results to form my predictions, the results from this experiment can lend some weight into the relationship between perception and production.

It is already known that the relationship between these two processes are more complex than previous studies have suggested (Baese-Berk & Samuel, 2016). This thesis hopes to lend weight to understanding how production and perception are linked in the context of native Standard Chinese speakers and tone three sandhi.

Gao et al., (2021)'s perception experiment suggested that based on their results, the surface form tone mapping plays a bigger role in compound verbs than in reduplicated verbs — in the context of sandhi. Therefore, if this were to overflow into production, perhaps the way native speakers pronounce these two categories are different tonal contours than initially predicted.

Based on the answer to the first research question, the difference was not strong, if present at all. There was a larger difference between T3 RED and T2 RED than there was between T3 COM and T3 RED — which brings up interesting ideas in and of itself to be discussed shortly.

These results bring up a division between perception and production. While there was suggested a difference in tone mapping from the perception experiment, the difference is not present in the produced material. This barrier provides critical knowledge that some aspects of perception are just in perception, similar to how there are underlying and surface forms of words. This conflict when it comes to mapping tones to compound and reduplicated verbs may only occur when native speakers have to read and perceive sounds, not when they read and produce it.

## 4.3 Unexpected Results

There were two unexpected results in the data that were not anticipated based on theories of the stimuli. One of these being Cluster 2's contour being a Tone 4 and the other being the slight difference in pronunciation of T2 RED in comparison to the other controls and stimuli categories.

### *4.3a Why Tone 4?*

The unexpected results of tone four is perplexing due to the fact that a) there are a lot of stimuli fitting into that category (46 out of initially 104) and b) there is limited theological explanation for why tone four should show up in the stimuli I have given my participants to produce.

There is, of course, Zhang and Lai (2010)'s production experiment concerning half-third sandhi. Half-third sandhi being when the underlying environment /T3 TX/ (X being any tone besides T3) will cut the initial T3 contour so that it is a low falling tone (much lower initial f0 than a T4 but for the sake of this thesis will be labeled as a 'weak T4') becoming [T4 TX]. While this could give reason for how a T3 could be seen as a T4, according to Zhang and Lai (2010), the environments in the stimuli should not have resulted in such.

T3 COM were the stimuli perfect for tone three sandhi /T3 T3/ becoming [T2 T3]. Furthermore T2 T3 COM were a control stimuli that Zhang and Lai (2010) specifically said it is the one environment that half-third sandhi was not productive.

And as for the reduplicated stimuli, it is possible for half-third sandhi to apply here as the reduplicated tone dropping makes the surface forms [T2 T0]. However, in Zhang and Lai (2010) they did not test neutral tones as the second syllable. Therefore, it is unclear whether it could be hypothesized that half-third sandhi occurs in such an environment.

However given that half-third sandhi has been studied in combination with tone three sandhi in the past, it could give reason for why tone four has appeared. It may not be a tone four at all but rather a weak tone four, as half-third sandhi suggests. Due to the normalized data and only plugging in expected T2's and T3's into the contour clustering software, the actual f0 of the data may only be shown in comparison to each other. No true T4 contours were plugged into the software, therefore while Cluster 2 may look like a drastic T4, it could be a much weaker T4 when compared to the full range of Standard Chinese tones and pitches.

If the contour was only a few stimuli, I would rule it as an outlier, but due to it being nearly half of the initially imputed data, there is something more going on here. Furthermore, the slight majority of data points from the T2 RED category in this T4 Cluster is another unexpected result.

### *4.3b Why T2 RED was different?*

T2 RED is the only category of stimuli whose majority fell into Cluster 2 — that looked like a T4. The reason for this could be due to the underlying forms of this category being different from every other category. All other categories of stimuli have, at one point or another, a T3 in their underlying form. T2 RED is the only category of stimuli that does not, going simply from /T2 T2/ to [T2 T0] without a T3 present at all.

The theorized surface form pitch contours for both T2 RED and T3 RED are identical [T2 T0], but the results show they are lumped into two different clusters. If these results are correct, the reason for the difference will likely be because of sandhi occurring before the tone dropping.

With the results at hand right now, it does appear that the inputs from T3 RED stimuli — whose theorized surface form is [T2 T0] — initial surface syllable [T2] does fit with every category that's surface form is [T2 T3] regardless of sandhi occurring (T3 COM) or not (T2 T3 COM). And notably does not fit with the category whose surface form is supposed to match it T2 RED.

Recall the transformations for T3 RED again as the first transformation before tone dropping is exactly [T2 T3]. The following table will illustrate this and show all stimuli categories underlying, medial transformation forms, and surface forms to further express how T2 RED is different than all of them.

**Table 8**

*Underlying to Surface Transformations for Stimuli*

| | Underlying Form | Tone Three Sandhi | Reduplicated Tone Dropping | Surface Form |
|---|---|---|---|---|
| **T3 COM** | /T3 T3/ | [T2 T3] | — | [T2 T3] |
| **T3 RED** | /T3 T3/ | [T2 T3] | [T2 T0] | [T2 T0] |
| **T2 RED** | /T2 T2/ | — | [T2 T0] | [T2 T0] |
| **T2 T3 COM** | /T2 T3/ | — | — | [T2 T3] |

In Table 8, you can see the highlighted forms that are all the forms that have a T3 in the second syllable of the form. Notice how every stimulus category has this *except* for T2 RED. This one difference could be the reason why T2 RED's contours were all different from T3 RED. Therefore it could be that native speakers *anticipate* a T3 to be in the second position in T3 RED — as they do in all the other categories in this cluster — T3 COM and T2 T3 COM — and pronounce it differently than when they anticipate a T2 in that same position in T2 RED.

Regardless, the only difference in all of the results is that T2 RED has the majority of their contours fitting in a different cluster than all other categories. None of T2 RED's underlying nor surface forms contain a T3, meanwhile all other categories do at one point or another. An anticipatory effect of pronouncing a T3 could be why T2 RED is pronounced differently than all other categories.

# 5 Conclusion

Overall, this thesis set out in an attempt to understand the underlying and surface forms of tone three sandhi in two morphological constructions: Compound and Reduplicated verbs. It further attempted to understand the complex relationship between perception and production through the avenue of converting a previous speech perception experiment into a production one. Through contour clustering software, produced speech of native speakers was analyzed for its tone contours of tone three sandhi. The results showed limited differences between the compound and reduplicated verbs. Concluding that, given the data, the difference in constructing morphology of compound and reduplicated verbs is not present in the produced tones. Furthermore, speech perceived differences between these two morphological constructions and sandhi does not have any evidence of produced differences identifying a barrier between perception and production.

# References


Audacity Team (n.d.) *Audacity(R)*. Available at: https://www.audacityteam.org (accessed June 2024).

Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36. https://doi.org/10.1016/j.jml.2015.10.008

Boersma, Paul & Weenink, David (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.13, retrieved 10 June 2024 from http://www.praat.org/

Chen, F., Mo, R., Huang, C.-R., & Chen, K.-J. (1992). Reduplication In Mandarin Chinese: Their Formation Rules, Syntactic Behavior And ICG Representation. In H.-H. Chen (Ed.), *Proceedings of Rocling V Computational Linguistics Conference V* (pp. 217–233). The Association for Computational Linguistics and Chinese Language Processing (ACLCLP). https://aclanthology.org/O92-1007

Chen, Matthew (2000). Tone Sandhi: Patterns across Chinese Dialects, Cambridge Studies in Linguistics 92 (Cambridge: Cambridge University Press)

Cheng, Chin-Chuan (1973). A Synchronic Phonology of Mandarin Chinese. Monographs on Linguistic Analysis 4 (The Hague: Mouton).

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Annual Review of Psychology*, *55*(Volume 55, 2004), 149–179. https://doi.org/10.1146/annurev.psych.55.090902.142028

Duanmu, S. (2005). *The Tone-Syntax Interface in Chinese: Some Recent Controversies*.

Duanmu, S. (2007). Introduction. In S. Duanmu (Ed.), *The Phonology of Standard Chinese* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780199215782.003.0001

Gao, F., Lyu, S., & Lin, C.-J. C. (2021). Processing Mandarin Tone 3 Sandhi at the Morphosyntactic Interface: Reduplication and Lexical Compounds. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.713665

Kaland, Constantijn. 2021. Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association*. doi:10.1017/S0025100321000049

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. Cognition, 38(3), 245–294. https://doi.org/10.1016/0010-0277(91)90008-R

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. *5th International Symposium on Tonal Aspects of Languages (TAL 2016)*, 33–37. https://doi.org/10.21437/TAL.2016-7

Mei, Tsu-Lin (1977). 'Tones and Tone Sandhi in 16th Century Mandarin', Journal of Chinese Linguistics, 5: 237–60.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173. https://doi.org/10.1016/j.cognition.2008.08.002

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. https://doi.org/10.1080/23273798.2014.942326

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Shen, J. (2022). A Study of the Neutral Tone in Standard Mandarin. *OALib*, *09*(08), 1–7. https://doi.org/10.4236/oalib.1109104

Shen, Jiong (1994). ‘Beijinghua shangsheng liandu de diaoxing zuhe he jiezou xingshi’ [Tonal patterns and rhythmic structure in successive third tones in the Beijing dialect], Zhongguo Yuwen, 1994.4: 274–81

Shih, Chilin (1986). ‘The prosodic domain of tone sandhi in Chinese’, doctoral dissertation, University of California, San Diego.

Shih, Chilin (1997). ‘Mandarin Third Tone Sandhi and Prosodic Structure’, in Wang Jialing and Norval Smith (eds.), Studies in Chinese Phonology (Berlin: Mouton de Gruyter), 81–123.

Wange, William S.-Y. and Kunk-Pu Li (1967). ‘Tone 3 in Pekinese’, Journal of the Speech and Hearing Research, 10.3: 629–36.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*(1), 61–83. https://doi.org/10.1006/jpho.1996.0034

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, *27*(1), 153–201.

Zhang, C., Xia, Q., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics*, *33*, 149–162. https://doi.org/10.1016/j.jneuroling.2014.07.002